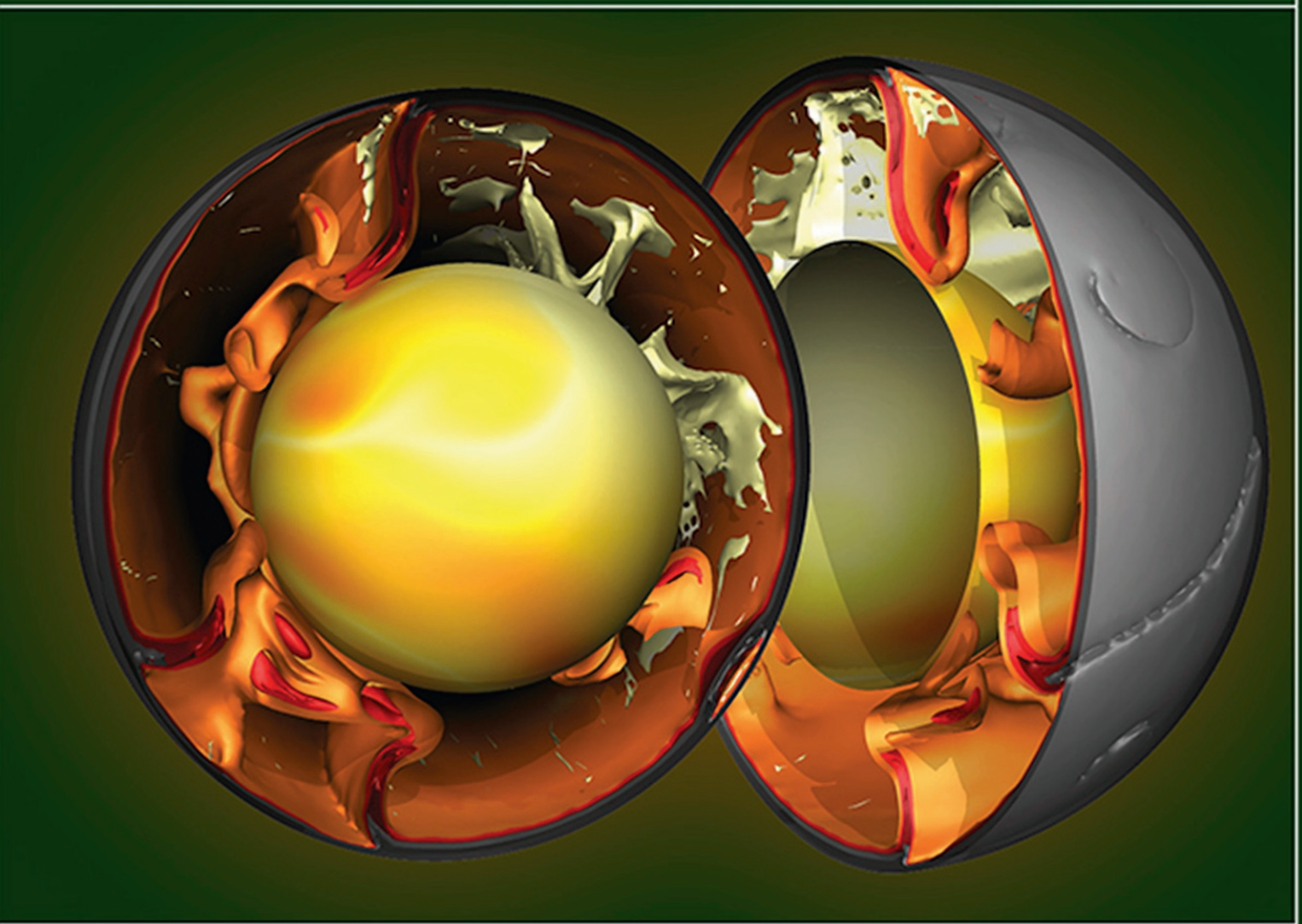
SECOND EDITION

TREATISE ON GEOPHYSICS

Volume 3

GEODESY



VOLUME EDITOR: THOMAS HERRING
EDITOR-IN-CHIEF: GERALD SCHUBERT

TREATISE ON GEOPHYSICS

SECOND EDITION

This page intentionally left blank

TREATISE ON GEOPHYSICS

SECOND EDITION

EDITOR-IN-CHIEF

Gerald Schubert

Distinguished Professor Emeritus, Department of Earth, Planetary and Space Sciences, University of California, Los Angeles

VOLUME 3

GEODESY

VOLUME EDITOR

Professor Thomas Herring

Department of Earth & Planetary Science, Massachusetts Institute of Technology

ELSEVIER AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
30 Corporate Drive, Suite 400, Burlington MA 01803, USA

Copyright © 2015 Elsevier B.V. unless otherwise stated. All rights reserved.

Cover image created for this work by Fabio Crameri, University College London.
The figure depicts the temperature field from a convection calculation using the code STAG-YY, written by Paul Tackley.

The following chapters are US Government works in the public domain and not subject to copyright:

1.11 Crust and Lithospheric Structure - Global Crustal Structure; 3.09 Earth Rotation Variations – Long Period; 4.07 The Role of Fault-Zone Drilling; 5.11 The Time-Averaged Field and Paleosecular Variation; 10.21 Exoplanetary Geophysics: An Emerging Discipline; 11.12 Tools and Techniques: Gravitational Method.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN (print): 978-0-444-53802-4

For information on all Elsevier publications
visit our website at http://store.elsevier.com/

Editorial & Production: Gemma Tomalin, Richard Berryman, Marise Willis, Michael Nicholls, Ellen Bishop
Cover design: Matthew Limbert

Printed and bound in the United Kingdom

15 16 17 18 19 10 9 8 7 6 5 4 3 2 1

CONTENTS

Preface *vii*

Editor-in-Chief *ix*

Contributors *xi*

Contents of all Volumes *xiii*

3.01 Geodesy: An Introduction and Overview 1
TA Herring

3.02 Potential Theory and the Static Gravity Field of the Earth 9
C Jekeli

3.03 Gravimetric Methods – Absolute and Relative Gravity Meter: Instruments Concepts and Implementation 37
T Niebauer

3.04 Superconducting Gravimetry 59
J Hinderer, D Crossley, and RJ Warburton

3.05 Gravimetric Methods – Satellite Altimeter Measurements 117
DP Chambers

3.06 Earth Tides 151
DC Agnew

3.07 Glacial Isostatic Adjustment and the Long-Wavelength Gravity Field 179
ME Tamisiea, JX Mitrovica, and K Latychev

3.08 Time-Variable Gravity from Satellites 193
JM Wahr

3.09 Earth Rotation Variations – Long Period 215
RS Gross

3.10 Earth Rotation Variations 263
V Dehant and PM Mathews

3.11 GPS and Space-Based Geodetic Methods 307
G Blewitt

3.12 Interferometric Synthetic Aperture Radar Geodesy 339
M Simons and PA Rosen

3.13 Geodetic Imaging Using Optical Systems 387
J-P Avouac and S Leprince

This page intentionally left blank

PREFACE

The *Treatise on Geophysics* is a comprehensive and in-depth study of the physics of the Earth and of terrestrial-like planets residing both within and outside our solar system. Its breadth and detail of coverage are beyond what any single geophysics text can provide. The first edition of the *Treatise on Geophysics* was published in 2007, nearly a decade ago. Of course, there is much progress in science in that length of time and the field of geophysics has grown rapidly and developed new lines of inquiry. Accordingly, we have brought the *Treatise on Geophysics* up-to-date with this second edition. The new edition will continue to provide students and professionals with fundamental and state-of-the-art discussion of all aspects of geophysics. In this new edition, the reader will find updates to all the chapters contained in the first edition and new chapters that discuss topics missing from the first edition. In a few instances, chapters from the first edition have simply been reprinted in order to avoid gaps in coverage. Chapters that emphasized fundamental physics often required little updating.

A highlight of the second edition is a new volume on Resources in the Near-Surface Earth. The new volume discusses the role of geophysics in the exploitation and conservation of natural resources and the assessment of degradation of natural systems by pollution. The near surface is a zone where humans and natural systems interact. Understanding the effects of human impacts is a challenge particularly for the long term.

There are four new chapters in the volume on Deep Earth Seismology. One of them describes progress in seismic instrumentation in the last 30 years in the era of broadband digital sensor and recording systems. Another one addresses a new approach for the determination of crust and uppermost mantle structure, based on ambient noise cross-correlations. This has revolutionized the field, since it allows the study of regional structure at high resolution in areas where there are no earthquakes. The third new chapter is devoted to recent advances in characterizing the lithosphere–asthenosphere boundary under both oceans and continents. Finally, the last new chapter discusses the morphology of subducted slabs in the Earth's transition zone.

The Mineral Physics volume has new chapters on dynamic (shock-wave) compression methods, and on the mineralogy of the upper mantle and the D″ layer focusing on post-perovskite, and extrasolar super-Earths.

New material in the Geodesy volume expands the temporal resolution and latency of geodetic measurements to high-rate, real-time measurements and increases the spatial resolution to meter size pixels with optical and radar imaging systems. Application areas are extended with new material on the uses of Earth gravity measurements.

The Earthquake Seismology volume has a new chapter on rock mechanics that is a comprehensive account of modern rock mechanics with emphasis on the variation of frictional properties of rocks during an earthquake cycle. Many chapters have considerable updates reflecting the lessons from significant recent earthquakes (e.g., the 2011 Tohoku-Oki earthquake and the 2013 Okhotsk deep earthquake), new observations on slow slips and triggered earthquakes, and developments of new methods (e.g., back projection method). Some chapters are updated (e.g., paleoseismology and strong ground motion) with substantial new material.

Chapters in the Geomagnetism volume have been updated to include recent advances in the acquisition of larger data sets and higher quality data and in the construction of more reliable and sophisticated models. Improvements in data have occurred in areas such as magnetic anomalies, paleointensities, and planetary magnetism. Improvements in models of the present day field include so-called comprehensive models and models that incorporate data assimilation.

The Crustal and Lithosphere Dynamics volume has new contributions on mountain building and landscape evolution.

The Mantle Dynamics volume contains new chapters on the generation of plate tectonics from mantle convection and the core–mantle boundary region.

The Core Dynamics volume features a comprehensive update of this rapidly changing subject, including new chapters on turbulence and waves in the fluid outer core.

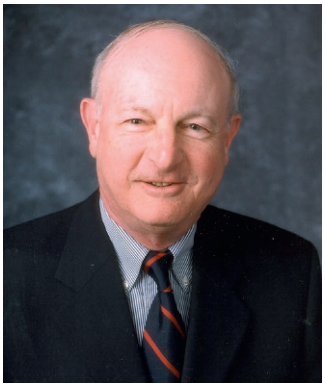
The volume on the Evolution of the Earth presents a new assessment on the role of volatiles in Earth evolution.

There are new chapters in the Physics of Terrestrial Planets and Moons volume on planetary atmospheres, planetary tectonics, extrasolar planets and asteroids and comets. The study of extrasolar planets has had explosive growth in the last decade and new spacecraft missions have significantly enhanced our knowledge of solar system planets and moons.

Finally, this is an opportunity for me to thank and acknowledge the editors of the individual volumes for their hard work, patience, dedication, and commitment in bringing this new edition of the Treatise to fruition. I am also indebted to Gemma Tomalin of Elsevier for her care and attention to the organization and production of the Treatise and to Deepa Arumugam for her careful coordination of the proofs from the copyeditor and typesetter. Of course, the authors of the individual chapters are the ones who are responsible for the excellence of the Treatise. We have all had some ups and downs in the process of producing the second edition of the Treatise but the final product is one that we can all view with pride.

Gerald Schubert
Editor-in-Chief

EDITOR-IN-CHIEF



Gerald Schubert is Distinguished Professor Emeritus in the Department of Earth, Planetary and Space Sciences at the University of California, Los Angeles. His research interests encompass the physics of the interiors and atmospheres of the Earth, the Moon, and the other moons and planets of the solar system. He is co-author with Donald Turcotte of *Geodynamics* (ed. 3, Cambridge University Press, 2014) and with Turcotte and Peter Olson of *Mantle Convection in the Earth and Planets* (Cambridge University Press, 2001). He is the author of over 540 research papers. He has participated in a number of NASA's planetary missions, including Apollo, Pioneer Venus, Magellan, and Galileo, and has been editor and editorial board member of many journals, including *Icarus*, *Journal of Geophysical Research*, *Geophysical Research Letters*, and *Annual Reviews of Earth and Planetary Sciences*. Professor Schubert is a Fellow of the American Geophysical Union and a recipient of the Union's James B. MacElwane medal and the Harry H. Hess medal. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences.

This page intentionally left blank

CONTRIBUTORS

DC Agnew
University of California, La Jolla, CA, USA

J-P Avouac
California Institute of Technology, Pasadena, CA, USA

G Blewitt
University of Nevada, Reno, NV, USA

DP Chambers
University of South Florida, St. Petersburg, FL, USA

D Crossley
St Louis University, Saint Louis, MO, USA

V Dehant
Royal Observatory of Belgium, Brussels, Belgium

RS Gross
California Institute of Technology, Pasadena, CA, USA

TA Herring
Massachusetts Institute of Technology, Cambridge, MA, USA

J Hinderer
Institut de Physique du Globe de Strasbourg, Strasbourg, France

C Jekeli
The Ohio State University, Columbus, OH, USA

K Latychev
University of Toronto, Toronto, Ontario, Canada

S Leprince
California Institute of Technology, Pasadena, CA, USA

PM Mathews
University of Madras, Chennai, India

JX Mitrovica
Harvard University, Cambridge, MA, USA

T Niebauer
Micro-g LaCoste, LLC, Lafayette, CO, USA

PA Rosen
California Institute of Technology, Pasadena, CA, USA

M Simons
California Institute of Technology, Pasadena, CA, USA

ME Tamisiea
National Oceanography Centre, Liverpool, UK

JM Wahr
University of Colorado, Boulder, CO, USA

RJ Warburton
GWR Instruments, San Diego, CA, USA

This page intentionally left blank

3.01 Geodesy: An Introduction and Overview

TA Herring, Massachusetts Institute of Technology, Cambridge, MA, USA

© 2015 Elsevier B.V. All rights reserved.

3.01.1	**Introduction**	1
3.01.2	**Coordinate Systems**	1
3.01.3	**Geodetic Methods**	3
3.01.3.1	Ground-Based Positioning Systems	3
3.01.3.2	Satellite Systems	6
3.01.3.3	Gravimeters	6
3.01.4	**Error Sources, Signals, and Noise**	8
3.01.5	**Conclusions**	8
References		8

3.01.1 Introduction

Modern geodesy as discussed in this volume started with the development of distance measurement using propagating electromagnetic signals and the launch of Earth-orbiting satellites. With these developments, space-based geodesy allowed global measurements of positions, changes in the rotation of the Earth, and the Earth's gravity field. These three areas, positioning, Earth rotation, and gravity field, are considered the three pillars of geodesy. The accuracy of current measurement systems allows time variations to be observed in all three areas. Also, the complexity of problems is such that each of the pillars interacts with each other and with many other branches of Earth science. This interaction is most apparent in the role that water plays in modern geodetic measurements. Every chapter in this volume mentions the role of water. It is critical because it can move rapidly and over large distances; it can exist in all three phases, gas, fluid, and solid; and modern geodetic methods are accurate enough that their measurements are sensitive to its effects. In its vapor form, its refractive properties delay microwave signals propagating through the Earth's atmosphere. For geodetic positioning, this is a noise source but it is a signal for meteorologic applications. As a liquid, it forms oceans that affect both the tidal signal and the rotation of the Earth. Also in liquid form, its mass changes the gravity field as it moves through the hydrologic cycle. In solid form, it has a gravitational and deformational signal that changes if melting the ice unloads the surface of the Earth. The interactions between the pillars include the elastic loading effects of changing mass loads that can be seen in the gravity field and in the positions of ground stations. The movement of water to and from the oceans can be seen with altimeter satellites whose orbital information is derived from measurements from ground stations whose positions are affected by the changing mass load. In modern, time-dependent geodetic data analysis, these interactions need to be accounted for. The common interface between the geodetic methods is the coordinate systems and reference frame used to analyze data.

Coordinate systems and the associated reference frames form a core theme in the chapters in this volume. Two other unifying themes are the measurement systems of geodesy that are used again throughout the volume and the interplay of errors in measurements, signals, and noise. In this overview, we examine these three themes.

3.01.2 Coordinate Systems

A common aspect of geodesy that permeates the literature of the subject is coordinate system definition. With the development of space-based methods that allow global measurements to be made, the subject has in some sense been greatly simplified in recent years while at the same time complicated by the increased accuracy needs. The simplification comes from being able to use a Cartesian coordinate system with an origin at the center of mass of the Earth and axes aligned in a well-defined manner to the outer surface of the Earth. The complications arise from needing a coordinate system definition that accounts for deformations from plate tectonics, tides, tectonic events, and other time-variable deformations. The measurement and understanding of these deformations are themes that are common to many of the chapters in this volume.

Historically, coordinate systems in geodesy are divided into two parts: horizontal coordinates such as latitude and longitude and a vertical coordinate called height. These two systems are fundamentally different in that the horizontal one is geometric and based on the direction of the normal to an ellipsoidal body that represents the average shape of the Earth. The height system, called orthometric height, is based on the gravitational potential field of the Earth. In this system, surfaces of constant height are equipotential surfaces, and hence, fluid will not flow along these surfaces. One special height surface is called the geoid and is associated with the surface representing mean sea level. In Chapters 3.02 and 3.05, the subtle problems with these definitions from defining an equipotential surface to the meaning of mean sea level in an ocean with currents are discussed. In conventional geodetic systems, there is a blend of potential-based and geometric systems. The determination of the geodetic latitude and longitude is complicated by measurements being made in the Earth's gravity field. Specifically with the development of electronic distance measurements, the distance measurement itself does not depend on the gravity field (except for small relativistic effects), but the

projection of the direct measurement to a horizontal measurement does depend on the slope of the measured line. The angle between measured direction and local vertical (which depends on the local direction of gravity) can be easily measured but is strictly not the correct measurement. The angle to the local normal to the ellipsoid is needed. The difference between these directions is called the 'deflection of the vertical' and as discussed in Jekeli can be determined from measurements of gravity and solving the appropriate boundary value problem. Deflection of vertical is also the difference between geodetic and astronomic latitude and longitude.

The blend of potential-based and geometry-based systems in geodetic coordinate systems is a consequence of the methods available for making measurements. The local nature of these measurements resulted in large differences among systems adopted by different countries. The reference ellipsoidal shape for the Earth could have differences in the semimajor axis of over a kilometer (e.g., Clarke 1866 6378206.0 m and Everest 6377276.0 m (Smith, 1996)), and it was not uncommon at boundaries of countries to have coordinate differences of several hundred meters. The advent of space-based measurements allows the determination of a best-fit ellipsoidal shape for the Earth based on global data. There are simple relationships between the gravity of the Earth, expressed in spherical harmonics, the moments of inertia of the Earth, and an appropriate flattening for the Earth. After the central force term in the gravity field, the flattening term is the largest, being at least 1000 times larger than any other term in the field. The leading terms in the gravitational potential field, V, of the Earth can be written in spherical harmonics as (e.g., Stacey, 1992)

$$V(r,\theta) = -\frac{GM}{r}\left(J_0 P_0 - J_1 \frac{a}{r} P_1(\cos\theta) - J_2 \left(\frac{a}{r}\right)^2 P_2(\cos\theta)\right) \quad [1]$$

where G is the gravitational constant; M is the adopted mass of the Earth; θ and r are colatitude and radial distances to the point where the potential is determined; a is the equatorial radius of the Earth; J_0, J_1, and J_2 are the coefficients of the gravity field; and P_0, P_1, and P_2 are Legendre functions. When the adopted mass of the Earth matches that of the Earth, the $J_0 P_0$ term is 1, and if the center of mass of the Earth corresponds to the origin of the coordinate system used to measure latitude, J_1 is zero. When only the largest terms in the gravity field are considered, the Earth is axially symmetrical, and thus, there is no longitude dependence. The second-degree harmonic term, J_2, is related to both the moments of inertia of the Earth, which will control the Earth's rotational behavior and the flattening of the shape of the Earth. The moments of inertia of the Earth can be expressed in terms of J_2 through MacCullagh's formula yielding

$$J_2 = (C - A)/(Ma^2) \quad [2]$$

where A and C are the minimum and maximum moments of inertia of the axially symmetrical Earth, respectively. By setting the potential to be the same at the equator and the pole, the semimajor and semiminor axes of an ellipsoid, a and c, that match these values allow the flattening of the ellipsoid to be derived from the gravity field. Neglecting higher-order terms in the gravity field, we have (Stacey, 1992)

$$f = \frac{a-c}{c} = \frac{C-A}{Ma^2}\left(\frac{a^2}{c^2} + \frac{c}{2a}\right) + \frac{1}{2}\frac{a^2 c\omega^2}{GM} \approx \frac{3}{2}J_2 + \frac{1}{2}m \quad [3]$$

where f is the flattening and the centrifugal component of acceleration from the rotation of the Earth (rotation rate ω) has been included. The ratio of the equatorial gravitational to rotational force is given by m. From eqns [1] to [3], we see the first-order relationships between the gravity field of the Earth, the moments of inertia of the Earth, and the shape of the Earth. With modern space-based measurements, all of these quantities can be measured with great precision. **Chapter 3.02** details the theory of gravitational potential, **Chapters 3.09** and **3.10** examine the rotational consequences of moments of inertia, and **Chapter 3.05** examines the detailed shape of the geoid based on altimeter measurements from space.

With the development of space-based geodetic systems, it became possible to define a purely geometric global coordinate system. These new systems are Cartesian and have their origin approximately at the center of mass of the Earth. The axes are aligned approximately along the rotation axis and the Greenwich meridian. The alignments here are only approximate because, with the accuracy of modern measurements, the center of mass and rotation axis move with respect the figure of the Earth. The rotation axis motions are large, of order 10 m over a year, and have been known about for over a century. The center-of-mass movements are much smaller, of order 1–2 cm, and have been measured for the past few decades. Even now, center-of-mass variations are not fully accounted for in modern coordinate systems. The current international terrestrial reference frames (ITRF) (Altamini, 2012; Altamini et al., 2002, 2011) have an origin at the center of the figure of the Earth. Mass movements on the surface of the Earth, mainly atmospheric and hydrospheric, result in the movement of the center of mass relative to the center of figure that needs to be accounted for in accurate representations of the gravity field. When center-of-figure reference frames are used, the gravity fields in these frames should have time-dependent first-degree spherical harmonic terms included. Currently, these terms need to be observationally determined and the routine production of such values is not yet available. **Chapter 3.08** discusses in detail the measurement of time-variable gravity. To further complicate accounting for center-of-mass movements, the mass movements that cause the center of mass to move, also typically load the surface, and hence deform the surface the Earth. The deconvolution of the effects of surface deformations affecting the center of figure and the mass contributions that affect center-of-mass positions is not trivial. This subject is explored in Lavallée and Blewitt (2002) and Blewitt (2003) and is discussed in **Chapter 3.11**.

The treatment of Earth rotation variations is commonplace in modern systems because the effects are so large. However, even with this topic, there are subtle problems associated with deformations of the Earth. The motion of the rotation axis with respect to the crust of the Earth is dominated by a resonance term called the Chandler wobble, named after the person to first observe the effect, and an annual signal due mainly to seasonal mass movements. The Chandler wobble has a period of about 430 days with the frequency related to the moments of inertia of the Earth, the elastic properties of the mantle, and the effects of the oceans and the fluid core. The annual signal and Chandler wobble cause motions of the pole with an amplitude of about 10 m. Superimposed on these large quasi-periodic signals are other broadband signals that arise from

movements in the atmosphere, oceans, and fluid core. The details of these motions are discussed in Chapter 3.09. In addition to these periodic-type variations, there is a secular drift of the pole due to changes of moments of inertia resulting from glacial isostatic adjustment. This area is discussed in Chapter 3.07.

In addition to changes in the position of the rotation axis with respect to the crust of the Earth, the rotation rate also varies with both a secular term, due to energy dissipation in the Earth–Moon system, and shorter-period variations due to momentum exchanges between the atmosphere, oceans, and fluid core and the solid Earth. For periods less than a few years, the atmosphere is the dominant source of changes in angular momentum. For longer-period terms, the fluid core is the most likely source. Rotation rate changes are discussed in Chapters 3.09 and 3.10.

The concept of rotation changes introduces another coordinate system that is needed in modern geodesy. This additional coordinate system is one that does not rotate in inertial space. Rotation rate variations are measured relative to this nonrotating system. Historically, this external, nonrotating frame has been defined by measurements to optical stars and to bodies in our solar system. Both of these systems have problems when very accurate measurements are needed. Easily seen optical stars are quite close to us and exhibit the so-called proper motions, meaning that they move relative to a nonrotating frame. Solar system measurements suffer from lack of accurate observations and their interpretation requires that the equations of motion of objects in the solar system contain all of the correct force models. So neither optical objects nor solar system dynamics provide a stable nonrotating coordinate system. Such a coordinate system is provided by extragalactic radio sources whose directions can be measured using very long-baseline interferometry (VLBI) (Ma et al., 1998). Changes in rotation rate are measured relative to this nonrotating system, and time as measured by the rotation of the Earth, Universal Time 1 or UT1, is the integration of the rotation rate.

With the definitions of an inertial coordinate system and one that is attached to the Earth, instantaneous rotation angles can be defined that rotate one system into the other. The spectrum of these rotation angles shows strong peaks in the nearly diurnal band and broad spectral response outside the diurnal band. Historically, the rotation vector has been divided into two parts. When the rotation axis is viewed from the rotating Earth, the nearly diurnal retrograde motion of the rotation axis represents a motion that is slow in inertial space and is called nutation. This motion has been known about since ancient Greek times. The word nutation is Greek for nodding. The nodding is caused by the time-dependent torques applied to the equatorial bulge of the Earth by the gravitational forces from the Moon, Sun, and planets. The average torque applied by these bodies causes a secular motion of the pole in inertial space called precession. The longer-period motion is called polar motion. In an approximate sense, nutation is caused by external torques applied to the Earth and pole motion due to mass and angular momentum redistributions in the Earth. The separation and interpretation of these motions are discussed in Chapters 3.09 and 3.10.

The final aspect of measuring changes in the rotation of the Earth is the definition of rotation angles. All points on the surface of the Earth move relative to the interior just as the interior itself moves due to mantle convection. To define rotation changes, the deformation component must be accounted for. For secular motions due to plate tectonics, the separation is made, making the average motion having no net rotation, the so-called no-net-rotation or NNR frame. The current definitions of rotation angles for the Earth as distributed by the International Earth Rotation Service use this definition. With modern geodetic accuracy, nonsecular motions of sites are now evident due mainly to loading effects. The gravitational consequences of the mass movements and loads are discussed in Chapter 3.08 but there are also consequences for Earth rotation measurements when small numbers of sites are used. For Earth rotation measurements using a large number of sites, some of these nonsecular site movements can be averaged. However, the spatial correlation of loading effects suggests that there is an upper limit to the number of sites that are needed to effectively measure Earth rotation changes and average out the loading effects and other nonsecular motions. Ultimately, in closed system analysis, nonsecular deformations would be accounted for, for example, from the analysis of gravity changes, before the rotation angles are computed.

3.01.3 Geodetic Methods

Modern geodetic methods have in recent decades pushed the precision of measurements and the completeness of models needed to explain these measurements. In geodesy, it is common to talk about the fractional precision of measurement. A 1 km distance measured with a precision of 1 mm is said to be precise to 1 part per million or 1 ppm. On large scales, modern geodetic measurements are precise to better than 1 part per billion (1 ppb) in many cases. For global height measurement, 1 ppb corresponds to 6 mm uncertainties in heights. For gravity measurements, 1 ppb is 10 nm s^{-2} or 1 μgal where gal is the cgs unit for gravity named after Galileo. The Earth's gravity is about 980 gals. There are a variety of geodetic systems that achieve these precisions. In this treatise volume, all the modern geodetic measurement methods are discussed. They fall into three basic classes: ground-based positioning systems that provide geometric positioning and tracking of external objects, satellite systems that sense the Earth's gravity field and/or make measurements directly from space, and ground-based instrumentation that measures the gravity field on the Earth's surface. The interpretation of geodetic measurements often requires inputs from all these systems.

3.01.3.1 Ground-Based Positioning Systems

Four major geodetic measurements fall into this category: VLBI, satellite laser ranging (SLR), global positioning system (GPS), and the Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) system. These systems use either microwave (VLBI, GPS, and DORIS) or optical (SLR) frequency signals to provide the carrier for their measurements. The basic measurement types include group and phase delay and Doppler shift measurements. Estimates of the positions and motions of locations on the Earth's surface obtained from the analysis of data from these systems are used in the formation of the ITRF. The latest version of the ITRF is ITRF2008 (Altimini et al., 2011; Altimini, 2012).

VLBI is a microwave-based measurement system that measures the difference in arrival times of incoherent signals from radio sources by cross correlation. Most commonly, the radio sources are extragalactic objects but beacons from satellites have also been used. Group and phase delays along with Doppler shift are measured. The phase delays are difficult to use in geodetic measurements with this system because they are measured modulo 2π, and reconstructing the number of cycles of phase is nontrivial in geodetic measurements. In the astronomical applications of VLBI, the use of the phase delays is common. Individual delay measurements have precisions of 1–10 mm. Generally, radio telescopes with diameters between 10 and 30 m are used for VLBI measurements, and because of the separation of the telescopes (thousands of kilometers), the delay measurements are made relative to independent hydrogen maser clocks at the observatories. The radio signals propagate through the Earth's atmosphere and are delayed through the refractive index of the atmosphere's gas constituents. VLBI measurements are made at two relatively high frequencies (usually ~2.3 and 8 GHz) and are delayed by the propagation in the ionosphere (removed with a dual-frequency correction since this medium is dispersive) and they are strongly affected by the dipole component of the refractivity of atmospheric water vapor. **Figure 1** shows the distribution of VLBI sites included in ITRF2008. Of the sites shown, about 40 stations are currently operating. VLBI requires coordinated measurements so that distant telescopes look at the same objects at the same time. Measurements with VLBI are usually 24 h duration observing sessions using four to eight radio telescopes around the world. VLBI activities are now coordinated through the International VLBI Service (Schlueter et al., 2002).

VLBI is the only modern technique with direct reference to a stable inertial reference frame. It uniquely contributes to maintaining time as measured by the rotation of the Earth and to monitoring the motion of the Earth's rotation axis in inertial space. These latter measurements provide unique insights into the interaction of the fluid-outer and solid-inner cores and the mantle of the Earth. This topic is discussed in detail in **Chapter 3.10**. VLBI also provides accurate position and atmospheric delay measurements.

SLR is an optical-based system that uses short pulsed laser and accurate timing equipment to measure the round trip flight time between a ground system and a satellite equipped with special corner cube retroreflectors. Some laser systems have enough power to make range measurements to corner cube arrays on the Moon. Specially equipped optical telescopes are used for these measurements. Measurement accuracies of 1–10 mm are common for modern systems. The optical frequencies used in these systems are not affected by the Earth's ionosphere, and the contribution from atmospheric water vapor is much less than for microwave systems (the optical frequencies are too high to efficiently excite the dipole water vapor resonance). **Figure 2** shows the distribution of SLR sites used in ITRF2008. Of these sites, about 35 are currently active. SLR measurements do not need to be closely coordinated, and hence, SLR stations can operate semiautonomously. Some stations operate continuously, while others operate on 8 h shifts. There are priories set for which of the current 26 satellites with corner cubes should be tracked when multiple satellites are visible. The coordination of SLR measurements is provided by the International Laser Ranging Service (Pearlman et al., 2002).

SLR provides measurements to about 26 orbiting spacecraft and the Moon. These measurements are used to determine the positions of the SLR sites and the orbits of the spacecraft. Perturbations to the spacecraft orbits are used to study the Earth's gravity field and its temporal variations. For geodetic positioning, the most commonly observed satellites are the pair of Laser Geodynamics Satellites (LAGEOS). These are high area-to-mass ratio satellites orbiting at about 6000 km altitude. At this altitude, the orbits are sensitive to the lower-degree terms in the Earth's gravity and measurements to LAGEOS provided the first measurements of secular changes in the lower-degree gravity field coefficients. The altitude is also high enough to be only slightly affected by atmospheric drag, although there are some not well-understood electromagnetic perturbations to the orbits. The high altitude of the LAGEOS

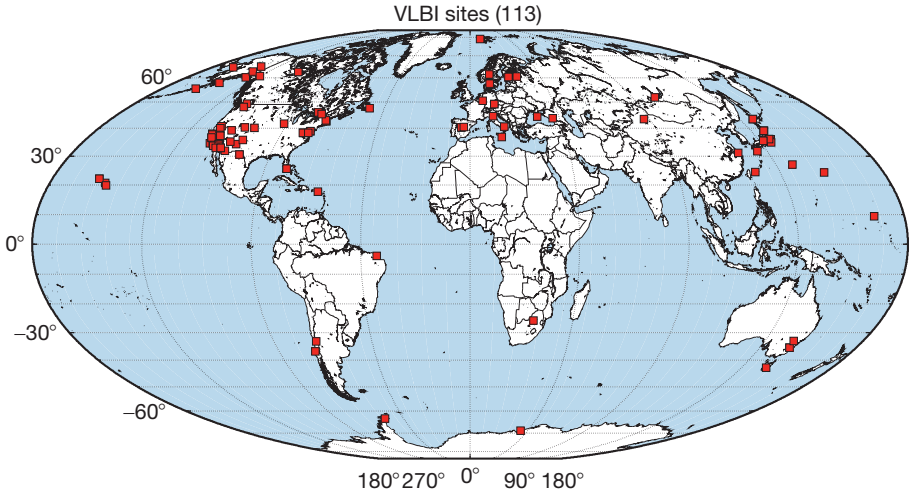


Figure 1 Locations of the 113 VLBI sites that are included in the ITRF2008 reference frame. Of the sites shown, about 40 are still actively making measurements. Measurements with VLBI are organized into sessions, usually of 24 h duration, using four to eight radio telescopes.

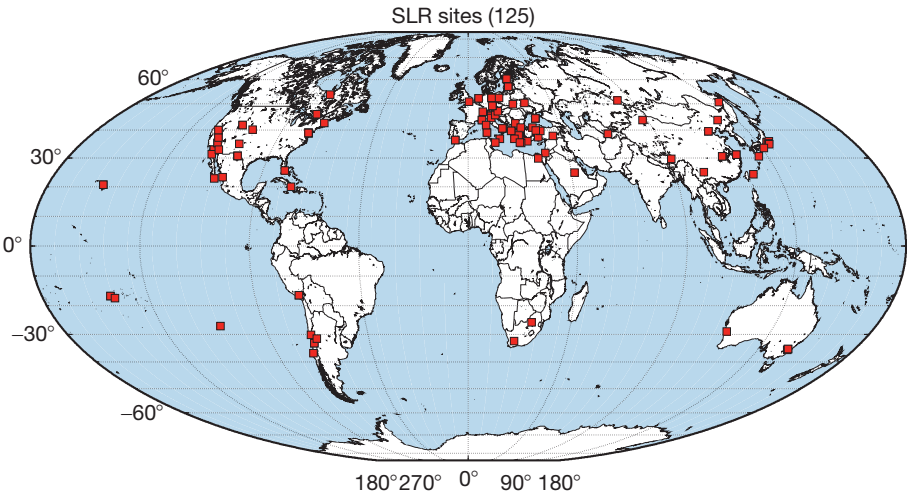


Figure 2 Locations of the 125 SLR sites that are included in the ITRF2008 reference frame. Of the sites shown, about 35 are still actively making measurements. These stations observe continuously with priorities, but not detailed scheduling, set by the international laser ranging service.

satellites also means that the center of mass of the Earth can be well determined from the measurements. The primary aim of much of the SLR tracking is precise orbit determination (POD). For the altimeter missions discussed in Chapter 3.05, the determination of orbits with sub-centimeter accuracy is critical.

The GPS is a microwave group and phase delay measurement system using relatively low L-band signals (~1.2 and 1.5 GHz). The system was initially designed as a military navigation system that used group delays. Time tags are encoded into the transmitted signal and a ground receiver can measure the difference in time between when a signal was transmitted from a satellite based on the satellite's clock and when it was received on the ground based on the ground receiver's clock. If the clocks were synchronized, the time difference would be a measure of range. In the navigation application, accurate clocks are placed on the satellites, and because the ground receivers can simultaneously make measurements to multiple satellites, the ground receiver's clock error can be estimated or eliminated by differencing measurements. The ability of GPS receivers to make measurements simultaneously to multiple satellites and the narrow bandwidth of the transmitted signals means relatively inexpensive electronics can be used to make very precise measurements. In the geodetic applications of GPS, the phase measurements are used, which is possible because multiple ground stations can see the same satellites, and by either estimation or differencing, the contributions of local oscillator phases in the receivers and transmitters can be eliminated. Phase measurements of range changes accurate to a few millimeters are possible. Dual-frequency measurements are used to eliminate the ionospheric delays, but because of the low frequency, the dual-frequency corrections are not always adequate. The second-order effects from the Earth's magnetic field need to be accounted for in the most accurate analyses. Continuous tracking of signals from the satellites allows the number of cycles in the phase to be counted, and only when the signal from the satellite is interrupted or when a satellite is first seen is it necessary to estimate the number of integer cycles in the phase. Figure 3 shows the GPS sites included in the ITRF2008 reference frame. We also show on this figure the additional 1000 GPS sites whose data are routinely available in the international GPS data archives. GPS stations track satellites continuously, are low power, and are suited to autonomous operation. Many thousands of GPS stations operate continuously around the world. The coordination of station standards and data analysis is through the International Global Navigation Satellite System Service (IGS) (Beutler et al, 1999). In addition to GPS, the IGS also includes measurements from receivers that track the Russian GLONASS satellites and in the future will include measurements from the European Galileo system. All these systems use similar frequency bands and their satellites are in medium Earth orbit with altitudes near 20 000 km (and 12 h orbital periods).

GPS and the other GNSS systems provide inexpensive, very precise positioning capabilities on both static and moving receivers. The systems are often deployed in dense networks to monitor tectonic deformations with continuous and occasional occupations. The large numbers of stations available provide robust measurements of polar motion and short-period variations in length of day. Dense networks of GPS stations are also used to monitor atmospheric water vapor, and in many cases, results are used in operational weather forecasting. The high data rates available with GPS with sampling frequencies up to 50 Hz make the system ideal for tracking fast-moving objects such as aircraft and for monitoring seismic events. Chapter 3.11 discusses many of the applications of GPS along with the other space geodetic positioning systems.

The DORIS system is also a microwave system that uses two-way Doppler tracking. Unlike the other geodetic systems, DORIS stations have active communication with the satellites being tracked. The satellites initiate communications with the ground stations. This system allows two-way communications that can cancel errors due to different frequencies in the oscillators in the spacecraft and ground stations. DORIS is primarily used for POD and plays a critical role in tracking satellites used for radar altimetry discussed in Chapter 3.05. Earth orientation measurements, center-of-mass variations, and station locations are also determined in the analysis of DORIS data. The activities of DORIS are coordinated under the International DORIS

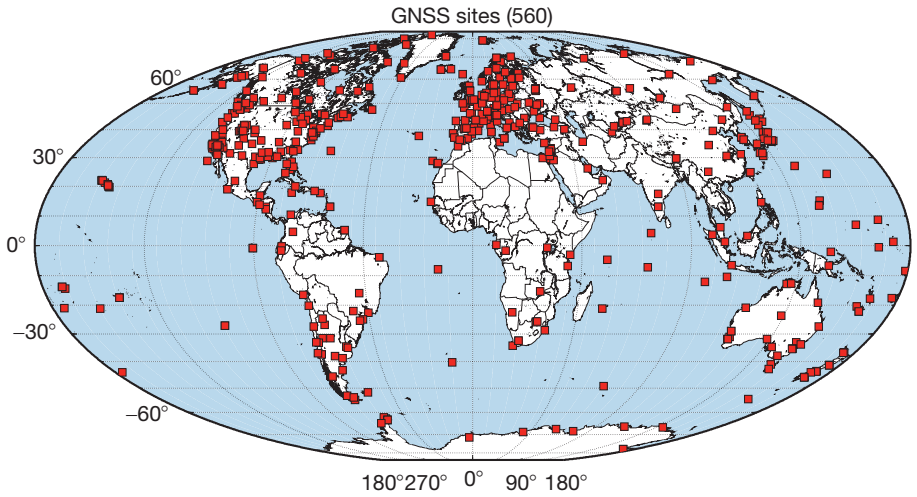


Figure 3 Locations of the 560 GPS sites that are included in the ITRF2008 reference frame (red squares). There are many additional sites whose data are made freely available through international data archives. These sites operate continuously.

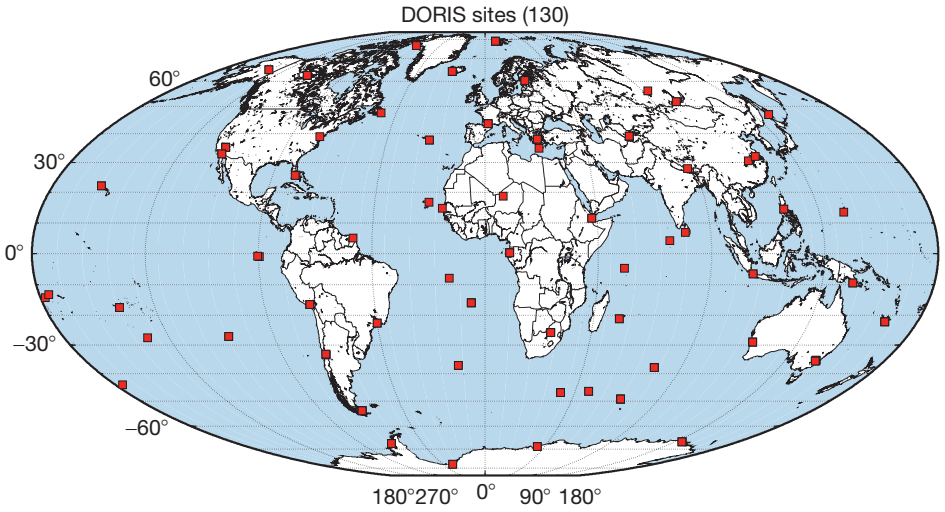


Figure 4 Locations of the 139 DORIS sites that are included in the ITRF2008 reference frame. These sites have active communication between the ground stations and the satellites that are tracked with DORIS.

service (Tavernier et al., 2006). In **Figure 4**, the DORIS sites used in the ITRF2008 are shown.

3.01.3.2 Satellite Systems

A wide variety of satellite systems are used in modern geodetic measurements. Those that are currently being tracked with laser ranging systems are listed in **Table 1**. The applications of satellites fall into different categories. Five of the entries are reflector arrays on the Moon, three from the Apollo program and two from the Russian Luna program. All of the Russian GLONASS satellites have corner cube arrays, as do two of the GPS satellites. The GPS arrays are used for accuracy evaluation. The LAGEOS, Starlette, and Stella spacecraft are passive, spherical bodies covered with corner cube reflectors and are primarily used for gravity field determination and positioning. The Jason-1 and ICESat satellites are microwave and laser altimeter satellites with Jason-1 used primarily for ocean studies and ICESat used primarily for ice sheet studies. **Chapter 3.05** discusses results from these and earlier altimeter satellites that are no longer operating. The CHAMP and GRACE satellites are used for gravity field studies with GRACE being a pair of satellites with accurate range and range-rate measurements between them. These measurements provide very accurate determinations of changes in the Earth's gravity field as detailed in **Chapter 3.08**. The ERS and Envisat satellites carry synthetic aperture radars, which when used in an interferometric mode, as discussed in **Chapter 3.12**, allow high-spatial-resolution maps of deformation fields or topography to be made.

3.01.3.3 Gravimeters

Gravimeter instrumentation for precise geodesy is classed as absolute gravimeters that normally use the acceleration of an

Table 1 Active geodetic satellites being tracked with SLR

Satellite name	*Satellite ID*	*Altitude (km)*	*Inclination (deg)*	*First tracked date*
Ajisai	8606101	1485	50	13 August 1986
Apollo 11 Sea of Tranquility	0000100	356 400	5	20 August 1969
Apollo 14 Fra Mauro	0000102	356 400	5	07 February 1971
Apollo 15 Hadley Rille	0000103	356 400	5	01 September 1971
Beacon-C	6503201	927	41	02 January 1976
COMPASS-G1	1000101	42 164	55.5	28 April 2012
COMPASS-I3	1101301	42 161	55.5	27 April 2012
COMPASS-I5	1107301	42 161	55.5	06 July 2012
COMPASS-M3	1201801	21 528	55.0	11 July 2012
CryoSat-2	1001301	720	92	20 April 2010
Envisat	0200901	800	98	10 April 2002
Etalon-1	8900103	19 105	65	26 January 1989
Etalon-2	8903903	19 135	65	13 July 1989
GIOVE-A	0505101	23 916	56	11 May 2006
GLONASS-102	0606201	19 140	65	04 May 2007
GLONASS-109	0706503	19 140	65	04 May 2007
GLONASS-110	0804601	19 140	65	10 December 2009
GLONASS-118	0907003	19 140	65	04 January 2010
GLONASS-129	1106402	19 140	65	02 January 2012
GLONASS-130	1107101	19 140	65	01 January 2012
GOCE	0901301	295	96.7	01 April 2009
GPS-36	9401601	20 030	55	21 April 1994
GRACE-A	0201201	485–500	89	18 March 2002
GRACE-B	0201202	485–500	89	18 March 2002
Galileo-101	1106001	23 220	56	29 November 2011
Galileo-102	1106002	23 220	56	29 November 2011
Galileo-103	1205501	23 220	56	07 November 2012
Galileo-104	1205502	23 220	56	07 November 2012
HY-2A	1104301	971	99.35	02 October 2011
IRNSS-1A	1303401	42 164	29	05 September 2013
Jason-2	0803201	1336	66	24 June 2008
KOMPSAT-5	1304201	550	97.6	09 September 2013
LAGEOS-1	7603901	5850	110	10 May 1976
LAGEOS-2	9207002	5625	53	24 October 1992
LARES	1200601	1450	69.5	17 February 2012
LRO-LR	0903101	~50 (lunar)	~90 (wrt lunar equator)	30 June 2009
Larets	0304206	691	98.204	04 November 2003
Luna 17 Sea of Rains	0000101	356 400	5	21 May 1975
Luna 21 Sea of Serenity	0000104	356 400	5	16 November 1973
QZS-1	1004501	32 000–40 000	45	11 September 2010
RadioAstron	1103701	500–350 000	51.4	15 November 2011
SARAL	1300901	814	98.55	04 March 2013
STPSat-2	1006201	650	72	02 April 2013
STSAT-2C	1300301	300–1500	80	29 March 2013
Starlette	7501001	815	50	03 January 1976
Stella	9306102	815	99	30 September 1993
TanDEM-X	1003001	514	97.44	21 June 2010
TerraSAR-X	0702601	514	97.44	16 June 2007

List available from and updated at http://ilrs.gsfc.nasa.gov/missions/satellite_missions/current_missions/index.html.
Source: International Laser Ranging Service.

object falling in vacuum or superconducting instruments that measure changes in gravity on a test mass that is cooled to very low temperatures to minimize the random accelerations from thermal noise. The applications and development of these classes of gravimeters are discussed in **Chapters 3.03** and **3.04**. The absolute gravimeters provide absolute measurements that are suitable for monitoring long-term changes in gravity due to either uplift or secular mass movements. These instruments are normally moved from place to place to make repeated measurements often over many years. The superconducting gravimeters provide continuous high-precision gravity measurements at observatory locations. They are suitable for tidal studies, particularly the effects of the fluid core on tidal amplitudes, seismic modes from large earthquakes, and atmospheric mass change excitations.

3.01.4 Error Sources, Signals, and Noise

Several common themes run through the chapters in this volume. The first of these is that geodesy is the combination of an observational science and a theoretical science. During various stages of its history, either new theories or new observations determine the direction of the science. During the last 20 years, observational accuracies have often been driving new theoretical developments. The other critical aspect of modern measurements is the errors in these measurements. At times, assessing error levels has been difficult, especially when accuracy is considered, because there have been no standards to compare with. Many of the chapters in this volume discuss the error spectra of geodetic measurements in the time and space domains. In many cases, a contribution that can be considered noise for one user may be a signal for others. The effects of water vapor and liquid water fall into the category of being both noise and signal.

An example of how theoretical and observational drivers have changed over the years is the study of nutation. For much of geodetic history, the motion of the rotation axis in space was computed theoretically from the ephemerides of the Moon and Sun and knowledge of the moments of inertia of the Earth. At the time that VLBI systems were being deployed worldwide and used on a regular basis, a new nutation theory was adopted by the International Astronomical Union (IAU) called the IAU 1980 theory of nutation. This new theory included effects of elasticity of the Earth (derived from seismic observations) and a fluid core whose shape was computed from the assumption of hydrostatic equilibrium throughout the Earth. Just a few years after the adoption of this theory, VLBI observations revealed discrepancies between the predictions of the theory and measurements, which resulted in a new assessment of the theory. Some parts of the assessment showed that the original theories were truncated at too coarse a level for the modern VLBI accuracies, and other parts showed that the theory needed to extend to include the effects of nonhydrostatic forces, magnetic coupling, and the presence of the solid-inner core. These developments are discussed in **Chapter 3.10**. **Chapter 3.04** discusses a very different measurement type that can be used to address the theory. Currently, there are theoretical predictions about the conductivity structure at the core–mantle boundary that generate results consistent with VLBI measurements that need to be assessed by an independent method.

3.01.5 Conclusions

Modern geodesy interacts with nearly all fields of Earth science, and the chapters in this volume show these interactions. The chapters progress through the different areas of geodesy starting with the theory of the Earth's gravity field, followed then by ground-based measurements of gravity and space-based altimetric measurements. The accuracy of the modern measurements allows temporal variations in gravity to be measured, and this topic is discussed in three chapters dealing with tidal variations in gravity, short-term variations that can now be measured from space, and long-period (secular) variations that are related to the viscoelastic response of the Earth to unloading of ice mass that occurred at the end of the last Pleistocene ice age. Of great interest currently is the separation of this secular signal from the signal due to current melting of ice sheets and glaciers. The combination of gravity measurements and surface uplift measurements shows promise for allowing this separation. The same mass movements that effect time-dependent gravity also change the rotation of Earth on many timescales. Two chapters address Earth rotation changes in inertial space and relative to the crust of the Earth. The volume concludes with the theory, instrumentation, and applications of modern position methods and surface deformation monitoring.

References

Altamimi Z, Collilieux X, and Métivier L (2011) ITRF2008: An improved solution of the International Terrestrial Reference Frame. *Journal of Geodesy* 85(8): 457–473. http://dx.doi.org/10.1007/s00190-011-0444-4.

Altamimi Z, Sillard P, and Boucher C (2002) ITRF2000: A new release of the International Terrestrial Reference Frame for earth science applications. *Journal of Geophysical Research* 107(B10): 2214. http://dx.doi.org/10.1029/2001JB000561.

Altimini Z (2012) http://itrf.ensg.ign.fr/ITRF_solutions/2008/more_ITRF2008.php.

Beutler G, Rothacher M, Schaer S, Springer TA, Kouba J, and Neilan RE (1999) The International GPS Service (IGS): An interdisciplinary service in support of earth sciences. *Advances in Space Research* 23(4): 631–635. http://igscb.jpl.nasa.gov.

Blewitt G (2003) Self-consistency in reference frames, geocenter definition, and surface loading of the solid Earth. *Journal of Geophysical Research* 108(B2): 2103. http://dx.doi.org/10.1029/2002JB002082.

Lavallée D and Blewitt G (2002) Degree-1 Earth deformation from very long baseline interferometry measurements. *Geophysical Research Letters* 29(20): 1967. http://dx.doi.org/10.1029/2002GL015883.

Ma C, Arias EF, Eubanks TM, et al. (1998) The International celestial reference frame as realized by very long baseline interferometry. *Astronomical Journal* 116: 516–546.

Pearlman MR, Degnan JJ, and Bosworth JM (2002) The International Laser Ranging Service. *Advances in Space Research* 30(2): 135–143. http://ilrs.gsfc.nasa.gov.

Schlueter W, Himwich E, Nothnagel A, Vandenberg N, and Whitney A (2002) IVS and its important role in the maintenance of the global reference systems. *Advances in Space Research* 30(2): 145–150. http://ivscc.gsfc.nasa.gov.

Smith JR (1996) *Introduction to Geodesy: The History and Concepts of Modern Geodesy*. *Wiley Series in Surveying and Boundary Control*, p. 224. New York: Wiley.

Stacey FD (1992) *Physics of the Earth*, 3rd edn. Brisbane: Brookfield Press, pp. 513.

Tavernier G, Fagard H, Feissel M, et al. (2006) The International DORIS Service: Genesis and early achievements. *Journal of Geodesy* 80(8–11): 403–428. http://ids.cls.fr/.

3.02 Potential Theory and the Static Gravity Field of the Earth

C Jekeli, The Ohio State University, Columbus, OH, USA

© 2015 Elsevier B.V. All rights reserved.

3.02.1	**Introduction**	9
3.02.1.1	Historical Notes	9
3.02.1.2	Coordinate Systems	10
3.02.1.3	Preliminary Definitions and Concepts	11
3.02.2	**Newton's Law of Gravitation**	12
3.02.3	**Boundary-Value Problems**	14
3.02.3.1	Green's Identities	15
3.02.3.2	Uniqueness Theorems	16
3.02.3.3	Solutions by Integral Equation	17
3.02.4	**Solutions to the Spherical BVP**	17
3.02.4.1	Spherical Harmonics and Green's Functions	17
3.02.4.2	Inverse Stokes' and Hotine Integrals	20
3.02.4.3	Vening-Meinesz Integral and Its Inverse	21
3.02.4.4	Concluding Remarks	22
3.02.5	**Low-Degree Harmonics: Interpretation and Reference**	23
3.02.5.1	Low-Degree Harmonics as Density Moments	23
3.02.5.2	Normal Ellipsoidal Field	24
3.02.6	**Methods of Determination**	26
3.02.6.1	Measurement Systems and Techniques	26
3.02.6.2	Models	30
3.02.7	**The Geoid and Heights**	31
References		33

3.02.1 Introduction

Classical gravitational potential theory has its roots in the late renaissance period when the position of the Earth in the cosmos was established on modern scientific, observation-based grounds. A study of Earth's gravitational field is a study of Earth's mass including its transport in time and its influence on near objects; and, fundamentally, it is a geodetic study of Earth's shape, which is described largely (70%) by the surface of the oceans. This initial section provides a historical backdrop to Newtonian potential theory and introduces some concepts in physical geodesy that set the stage for later formulations.

3.02.1.1 Historical Notes

Gravitation is a physical phenomenon so pervasive and incidental that humankind generally has taken it for granted with scarcely a second thought. The Greek philosopher Aristotle (384–322 BC) allowed no more than to assert that gravitation is a natural property of material things that causes them to fall (or rise, in the case of some gases) and the more material, the greater the tendency to do so. It was enough of a self-evident explanation that it was not yet to receive the scrutiny of the scientific method, the beginnings of which, ironically, are credited to Aristotle. Almost two thousand years later, Galileo Galilei (1564–1642) finally took up the challenge to understand gravitation through observation and scientific investigation. His experimentally derived law of falling bodies corrected the Aristotelian view and divorced the effect of gravitation from the mass of the falling object – all bodies fall with the same acceleration. This truly monumental contribution to physics was, however, only a local explanation of how bodies behaved under gravitational influence. Johannes Kepler's (1571–1630) observations of planetary orbits pointed to other types of laws, principally an inverse-square law according to which bodies are attracted by forces that vary with the inverse square of distance. The genius of Isaac Newton (1642–1727) brought it all together in his *Philosophiæ Naturalis Principia Mathematica* of 1687 with a single and simple all-embracing law that in one bold stroke explained the dynamics of the entire universe (today, there is more to understanding the dynamics of the cosmos, but Newton's law remarkably holds its own). The mass of a body was again an essential aspect, not as a self-attribute as Aristotle had implied, but as the source of attraction for other bodies: each material body attracts every other material body according to a very specific rule (Newton's law of gravitation; see **Section 3.02.2**). Newton regretted that he could not explain exactly why mass has this property (as one still yearns to know today within the standard models of particle and quantum theories). Even Albert Einstein (1879–1955) in developing his general theory of relativity (i.e., the theory of gravitation) could only improve on Newton's theory by incorporating and explaining action at a distance (gravitational force acts with the speed of light as a fundamental tenet of the theory). Of course, Einstein did much more by reinterpreting gravitational attraction as a warping of space and time, which then allows consideration of gravitation effects at high velocity and due to great quantities of mass (e.g., black holes). However,

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00056-7

what actually mediates the gravitational effect still intensely occupies modern physicists and cosmologists.

Gravitation since its early scientific formulation initially belonged to the domain of astronomers, at least as far as the observable universe was concerned. Theory successfully predicted the observed perturbations of planetary orbits and even the location of previously unknown new planets (Neptune's discovery in 1846 based on calculations motivated by observed perturbations in Uranus' orbit was a major triumph for Newton's law). However, it was also ascertained by measurement that gravitational acceleration varies on Earth's surface, with respect to both altitude and latitude. Newton's law of gravitation again provided the backdrop for the variations observed with pendulums. An early achievement for his theory came when he successfully predicted the polar flattening in Earth's shape on the basis of hydrostatic equilibrium, which was confirmed finally (after some controversy) with geodetic measurements of long triangulated arcs in 1736–37 by Pierre-Louis Moreau de Maupertuis and Alexis-Claude Clairaut and independently by Pierre Bouguer, Louis Godin, and Charles Marie de La Condamine in 1739–1740. Gravitation thus already played an early role in geodesy, the science of determining the size and shape of the Earth, and came to dominate it through promulgation in large part by the father of modern geodesy, Friedrich R. Helmert (1843–1917).

Terrestrial gravitation through the twentieth century was considered a geodetic area of research, although, of course, its geophysical exploits should not be overlooked. But the advancement in modeling accuracy and global application was promoted mainly by geodesists who needed a well-defined reference for heights (a level surface) and whose astronomic observations of latitude and longitude needed to be corrected for the irregular direction of gravitation. Today, the modern view of a height reference for ordinary positioning is changing to that of a geometric, mathematical surface (an ellipsoid), and three-dimensional coordinates (latitude, longitude, and height) of points on the Earth's surface are readily obtained geometrically by ranging to the satellites of the Global Positioning System (GPS) or other Global Navigation System of Satellites (GNSSs).

The requirements of gravitation for GPS (GNSS) orbit determination within an Earth-centered coordinate system are now largely met with existing models. Improvements in gravitational models are motivated in geodesy primarily for rapid determination of traditional heights with respect to a level surface (the geoid). These heights, for example, the orthometric heights, in this sense then become derived attributes of points, rather than their cardinal components. They are essential for any hydrologic applications since they determine the flow of water. The relatively expensive and time-consuming traditional determination of heights by spirit leveling would be replaced by the combination of GPS (GNSS) heights above the ellipsoid and a model of the geoid (see also **Section 3.02.7**). Indeed, both Canada in 2013 and the United States (within a decade) plan to replace their vertical height data with a geoid determined strictly from a geopotential model (Huang and Véronneau, 2013; NGS, 2012).

Navigation and guidance exemplify a further specific niche where gravitation continues to find important relevance. While GPS dominates also this field, those vehicles requiring completely autonomous, self-contained systems must rely on inertial instruments (accelerometers and gyroscopes). These do not sense gravitation (see **Section 3.02.6.1**), yet gravitation contributes to the total definition of the vehicle trajectory, and thus, the output of inertial navigation systems must be compensated for the effect of gravitation. By far, the greatest emphasis in gravitation, however, has shifted to the Earth sciences, where detailed knowledge of the configuration of masses (the solid, liquid, and atmospheric components) and their transport and motion leads to improved understanding of the Earth systems (climate, hydrologic cycle, and tectonics) and their interactions with life. Oceanography, in particular, also requires a detailed knowledge of a level surface, the geoid, to model surface currents using satellite altimetry. Clearly, there is an essential temporal component in these studies, and, indeed, the temporal gravitational field holds center stage in many new investigations. Moreover, Earth's dynamic behavior influences point coordinates and Earth-fixed coordinate frames (Plag and Pearlman, 2009), and we come back to fundamental geodetic concerns in which the gravitational field plays an essential role!

This chapter deals with the static gravitational field. The theory of the potential from the classical Newtonian standpoint provides the foundation for modeling the field and thus deserves the focus of the exposition. The temporal part is a natural extension that is readily achieved with the addition of the time variable (no new laws are needed, if we neglect general relativistic effects) and will not be expounded here. We are primarily concerned with gravitation on and external to the solid and liquid Earth since this is the domain of most applications. The internal field can also be modeled for specialized purposes (such as submarine navigation), but internal geophysical modeling, for example, is done usually in terms of the sources (mass density distribution), rather than the resulting field.

3.02.1.2 Coordinate Systems

Modeling the Earth's gravitational field depends on the choice of a coordinate system. Customarily, owing to the Earth's general shape, a spherical polar coordinate system serves for most applications, and virtually all global models use these coordinates. However, the Earth is slightly flattened at the poles, and a spheroidal coordinate system has also been advocated for some near Earth applications. We note that the geodetic coordinates associated with a geodetic datum (based on an ellipsoid) are never used in a foundational sense to model the gravitational field since they do not admit to a separation-of-variables solution of Laplace's differential equation (**Section 3.02.4.1**).

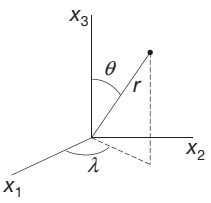


Figure 1 Spherical polar coordinates.

Spherical polar coordinates, described with the aid of Figure 1, comprise the spherical colatitude, θ; the longitude, λ; and the radial distance, r. Their relation to Cartesian coordinates is

$$\begin{aligned} x_1 &= r\sin\theta\cos\lambda \\ x_2 &= r\sin\theta\sin\lambda \\ x_3 &= r\cos\theta \end{aligned} \qquad [1]$$

Considering Earth's polar flattening, a better approximation than a sphere of its mostly oceanic surface is an ellipsoid of revolution or spheroid (the term 'ellipsoid' is used here as conventionally used in physical geodesy). Such a surface is generated by rotating an ellipse about its minor axis, which is also the polar axis. The two focal points of the best-fitting, Earth-centered ellipsoid are located in the equator about $E = 522$ km from the center of the Earth. A given linear eccentricity, E, defines the corresponding set of coordinates, as described in Figure 2. The longitude is the same as in the spherical case. The colatitude, δ, is the complement of the so-called reduced latitude; and the distance coordinate, u, is the semiminor axis of the confocal ellipsoid through the point in question. We call (δ,λ,u) *spheroidal* coordinates; they are also known as Jacobi ellipsoidal coordinates, or sometimes simply ellipsoidal coordinates. Their relation to Cartesian coordinates is given by

$$\begin{aligned} x_1 &= \sqrt{u^2+E^2}\sin\delta\cos\lambda \\ x_2 &= \sqrt{u^2+E^2}\sin\delta\sin\lambda \\ x_3 &= u\cos\delta \end{aligned} \qquad [2]$$

Points on Earth's best-fitting ellipsoid all have $u = b$; and all surfaces, $u =$ constant, are confocal ellipsoids (the analogy to the spherical case, when $E = 0$, should be evident). We note that these coordinates are not the same as Lamé's ellipsoidal coordinates for triaxial ellipsoids (Hobson, 1965).

3.02.1.3 Preliminary Definitions and Concepts

The gravitational potential, V, of the Earth is generated by its total mass density distribution. For applications on the Earth's surface, it is useful to include a potential-analogous function, ϕ, associated with the centrifugal acceleration due to Earth's rotation. The sum, $W = V + \phi$, is then known, in geodetic terminology, as the gravity potential, distinct from gravitational potential. The centrifugal potential, however, is not a Newtonian potential and is introduced for convenience only. It is further advantageous to define a relatively simple reference potential, or normal potential, that accounts for the bulk of the gravity potential (Section 3.02.5.2). The normal gravity potential, U, is defined as a gravity potential associated with a best-fitting ellipsoid, the normal ellipsoid, which rotates with the Earth and is also a surface of constant potential in this field. The difference between the actual and the normal gravity potentials is known as the disturbing potential: $T = W - U$; it thus excludes the centrifugal potential. The normal gravity potential accounts for approximately 99.9995% of the total potential.

The gradient of the potential is an acceleration, gravity, or gravitational acceleration, depending on whether or not it includes the centrifugal acceleration. Normal gravity, γ, comprises generally about 99.995% of the total gravity, g, although the difference in magnitudes, the gravity disturbance, δg, can be as large as several parts in 10^4. A special kind of difference, called the gravity anomaly, Δg, is defined as the difference between gravity at a point, P, and normal gravity at a corresponding point, Q, where $W_P = U_Q$ and P and Q are on the same perpendicular to the normal ellipsoid.

The surface of constant gravity potential, W_0, that closely approximates mean sea level is known as the geoid. If the constant normal gravity potential, U_0, on the normal ellipsoid is equal to the constant gravity potential of the geoid, then the gravity anomaly on the geoid is the difference between gravity on the geoid and normal gravity on the ellipsoid at respective points, P_0 and Q_0, sharing the same perpendicular to the ellipsoid. The separation between the geoid and the ellipsoid is known as the geoid undulation, N, or also the geoid height (Figure 3). A simple Taylor expansion of the normal gravity potential along the ellipsoid perpendicular yields the following important formula:

$$N = \frac{T}{\gamma} \qquad [3]$$

This is Bruns' equation (Hofmann-Wellenhof and Moritz, 2005), which is accurate to a few millimeters in N and which can be extended to $N = T/\gamma - (W_0 - U_0)/\gamma$ for the general case, $W_0 \neq U_0$. The gravity anomaly (on the geoid) is the gravity disturbance corrected for the evaluation of normal gravity on the ellipsoid instead of the geoid. This correction is $N\partial\gamma/\partial h = (\partial\gamma/\partial h)(T/\gamma)$, where h is height along the ellipsoid perpendicular. We have approximately $\delta g = -\partial T/\partial h$ and hence

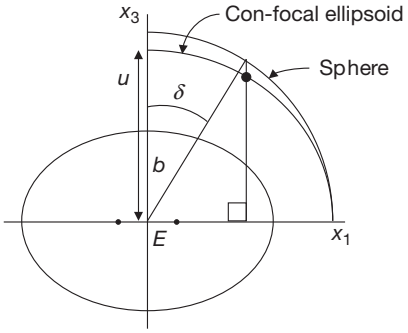


Figure 2 Ellipsoidal coordinates.

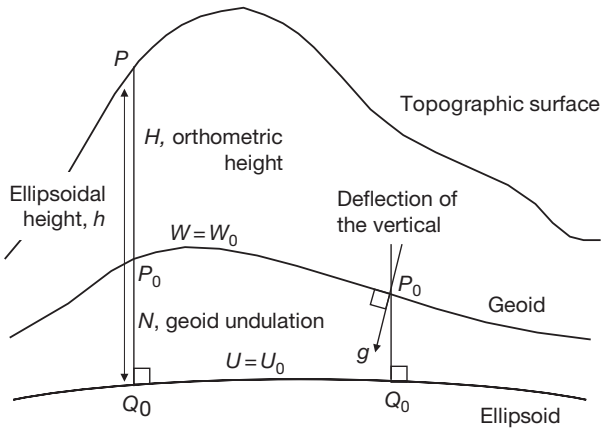


Figure 3 Relative geometry of geoid and ellipsoid.

$$\Delta g = -\frac{\partial T}{\partial h} + \frac{1}{\gamma}\frac{\partial \gamma}{\partial h}T \qquad [4]$$

which is known as the fundamental equation of physical geodesy (Hofmann-Wellenhof and Moritz, 2005).

The slope of the geoid with respect to the ellipsoid is also the angle between the corresponding perpendiculars to these surfaces. This angle is known as the deflection of the vertical, that is, the deflection of the plumb line (perpendicular to the geoid) relative to the perpendicular to the normal ellipsoid. The deflection angle has components, ξ and η, respectively, in the north and east directions. The spherical approximations to the gravity disturbance, the gravity anomaly, and the deflection of the vertical are given by (Hofmann-Wellenhof and Moritz, 2005)

$$\delta g = -\frac{\partial T}{\partial r}, \quad \Delta g = -\frac{\partial T}{\partial r} - \frac{2}{r}T,$$

$$\xi = \frac{1}{\gamma}\frac{\partial T}{r\partial \theta}, \quad \eta = -\frac{1}{\gamma}\frac{\partial T}{r\sin\theta\ \partial \lambda}, \qquad [5]$$

where the signs on the derivatives are a matter of convention.

3.02.2 Newton's Law of Gravitation

In its original form, Newton's law of gravitation applies only to idealized point masses. It describes the force of attraction, $\boldsymbol{F}$, experienced by two such solitary masses as being proportional to the product of the masses, m_1 and m_2; inversely proportional to the squared distance, ℓ^2, between them; and directed along the line joining them:

$$\boldsymbol{F} = G\frac{m_1 m_2}{\ell^2}\boldsymbol{n} \qquad [6]$$

where G is a constant, known as Newton's gravitational constant, that takes care of the units between the left- and right-hand sides of the equation; it can be determined by experiment, and the current internationally (2010 CODATA) accepted value is (Mohr et al., 2012)

$$G = (6.67384 \pm 0.00012) \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}} \qquad [7]$$

The unit vector $\boldsymbol{n}$ in eqn [6] is directed from either point mass to the other, and thus, the gravitational force is attractive and applies equally to one mass as the other. Newton's law of gravitation is universal as far as we know, breaking down when action at a distance cannot be assumed instantaneous and when very large velocities and masses are involved. Those cases must rely on Einstein's more comprehensive theory of general relativity, which describes gravitation as a characteristic curvature of the space–time continuum. In most geophysical and geodetic applications, the Newtonian theory is entirely adequate and it is the basis of the present exposition.

We can ascribe a gravitational acceleration to the gravitational force, which represents the acceleration that one mass undergoes due to the gravitational attraction of the other. Specifically, from the law of gravitation, we have (for point masses) the gravitational acceleration of m_1 due to the gravitational attraction of m_2:

$$\boldsymbol{g} = G\frac{m_2}{\ell^2}\boldsymbol{n} \qquad [8]$$

where the vector $\boldsymbol{g}$ is independent of the mass, m_1, of the body being accelerated (which Galileo found by experiment).

By the law of superposition, the gravitational force, or the gravitational acceleration, due to many point masses is the vector sum of the forces or accelerations generated by the individual point masses. Manipulating vectors in this way are certainly feasible, but fortunately, a more appropriate concept of gravitation as a scalar field simplifies the treatment of arbitrary mass distributions.

This more modern view of gravitation (already adopted by Gauss (1777–1855) and Green (1793–1841)) holds that it is a field having a gravitational potential. The concept of the gravitational field was fully developed by Lagrange (1736–1813) and Laplace (1749–1827), among others (Kellogg, 1953); and the potential, V, of the gravitational field is defined in terms of the gravitational acceleration, $\boldsymbol{g}$, that a test particle would undergo in the field according to the equation

$$\nabla V = \boldsymbol{g} \qquad [9]$$

where ∇ is the gradient operator (a vector of operators). Further elucidation of gravitation as a field grew from Einstein's attempt to incorporate gravitation into his special theory of relativity where no reference frame has special significance above all others. It was necessary to consider that gravitational force is not a real force (i.e., it is not an applied force, like friction or propulsion) – rather, it is known as a kinematic force, that is, one whose action is proportional to the mass on which it acts (like the centrifugal force) (Martin, 1988).

Though continuing with the classical Newtonian potential, we specifically interpret gravitation as an acceleration different from the acceleration induced by real, applied forces. This becomes especially important when considering the measurement of gravitation (**Section 3.02.6.1**). The gravitational potential, V, is a *scalar* function, and as defined here, V is derived directly on the basis of Newton's law of gravitation. To make it completely consistent with this law and thus declare it a Newtonian potential, we must impose the following conditions:

$$\lim_{\ell\to\infty} \ell V = Gm \quad \text{and} \quad \lim_{\ell\to\infty} V = 0 \qquad [10]$$

where m is the attracting mass and we say that the potential is regular at infinity. It is easy to show that the gravitational potential at any point in space due to a point mass, in order to satisfy eqns [8]–[10], must be

$$V = \frac{Gm}{\ell} \qquad [11]$$

where, again, ℓ is the distance between the mass and the point where the potential is expressed. Note that V for $\ell = 0$ does not exist in this case; that is, the field of a point mass has a singularity. We use here the convention that the potential is always positive (in contrast to physics, where it is usually defined to be negative, conceptually closer to potential energy).

Applying the law of superposition, the gravitational potential of many point masses is the sum of the potentials of the individual points (see **Figure 4**):

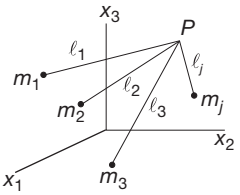


Figure 4 Discrete set of mass points (superposition principle).

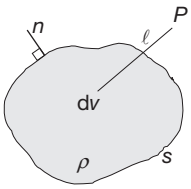


Figure 5 Continuous density distribution.

$$V_P = G\sum_j \frac{m_j}{\ell_j} \qquad [12]$$

For infinitely many points in a closed bounded region with infinitesimally small masses, dm, the summation in eqn [12] changes to an integration

$$V_P = G\int_{\text{mass}} \frac{dm}{\ell} \qquad [13]$$

or changing variables (i.e., units), $dm = \rho dv$, where ρ represents density (mass per volume) and dv is a volume element, we have (Figure 5)

$$V_P = G\iiint_{\text{volume}} \frac{\rho}{\ell} dv \qquad [14]$$

In eqn [14], ℓ is the distance between the evaluation point, P, and the point of integration. In spherical polar coordinates (Section 3.02.1.2), these points are (θ,λ,r) and (θ',λ',r'), respectively. The volume element in this case is given by $dv = r'^2 \sin\theta' \, d\lambda' \, d\theta' \, dr'$. V and its first derivatives are continuous everywhere – even in the case that P is on the bounding surface or inside the mass distribution, where there is the apparent singularity at $\ell = 0$. In this case, by changing to a coordinate system whose origin is at P, the volume element becomes $dv = \ell^2 \sin\psi \, d\alpha \, d\psi \, d\ell$ (for some different colatitude and longitude ψ and α); and, clearly, the singularity disappears – the integral is said to be weakly singular.

Suppose that the density distribution over the volume depends only on radial distance (from the center of mass): $\rho = \rho(r')$, and that P is an exterior evaluation point. The surface bounding the masses necessarily is a sphere (say, of radius, R), and because of the spherically symmetrical density, we may choose the integration coordinate system so that the polar axis passes through P. Then,

$$\ell = \sqrt{r'^2 + r^2 - 2r'r\cos\theta'}, \quad d\ell = \frac{1}{\ell} r' r \sin\theta' \, d\theta' \qquad [15]$$

It is easy to show that with this change of variables (from θ to ℓ), the integral [14] becomes simply

$$V(\theta,\lambda,r) = \frac{GM}{r}, \quad r \geq R \qquad [16]$$

where M is the total mass bounded by the sphere. This shows that to a very good approximation, the external gravitational potential of a planet such as the Earth (with concentrically layered density) is the same as that of a point mass.

The gravitational potential that is generated by other homogeneous bodies has been formulated in various ways, usually in terms of first and second spatial derivatives, conforming to gravitational measurements (Section 3.02.6), and in Cartesian coordinates. For example, Hofmann-Wellenhof and Moritz (2005) derived the expressions for the potential and its vertical derivative due to a cylinder along its axis. The following are expressions for the first and second derivatives of the potential due to a right rectangular prism with sides, $a_1 \leq x_1' \leq a_2$, $b_1 \leq x_2' \leq b_2$, and $c_1 \leq x_3' \leq c_2$, and valid in the exterior space. For $g_j = \partial V/\partial x_j$,

$$g_j = -G\rho\left(\left(x_k - x_k'\right)\ln\left(x_\ell - x_\ell' + r\right) + \left(x_\ell - x_\ell'\right)\ln\left(x_k - x_k' + r\right) - \left(x_j - x_j'\right)\tan^{-1}\frac{\left(x_k - x_k'\right)\left(x_\ell - x_\ell'\right)}{\left(x_j - x_j'\right)r}\right)\Bigg|_{x_1'=a_1}^{a_2}\Bigg|_{x_2'=b_1}^{b_2}\Bigg|_{x_3'=c_1}^{c_2} \qquad [17]$$

and for $\Gamma_{jk} = \partial^2 V/(\partial x_j \partial x_k)$,

$$\Gamma_{jj} = G\rho\tan^{-1}\frac{\left(x_k - x_k'\right)\left(x_\ell - x_\ell'\right)}{\left(x_j - x_j'\right)r}\Bigg|_{x_1'=a_1}^{a_2}\Bigg|_{x_2'=b_1}^{b_2}\Bigg|_{x_3'=c_1}^{c_2} \qquad [18]$$

$$\Gamma_{jk} = -G\rho\ln\left(x_\ell - x_\ell' + r\right)\Bigg|_{x_1'=a_1}^{a_2}\Bigg|_{x_2'=b_1}^{b_2}\Bigg|_{x_3'=c_1}^{c_2} \qquad [19]$$

where (x_1,x_2,x_3) is the point of computation and (j,k,ℓ) is a cyclic permutation of $(1,2,3)$. Formulas for general polyhedra may be found, for example, in Petrovic (1996) and Tsoulis and Petrovic (2001).

Besides volumetric mass (density) distributions, it is of interest to consider surface distributions. Imagine an infinitesimally thin layer of mass on a surface, s, where the units of density in this case are those of mass per area. Then, analogous to eqn [14], the potential is

$$V_P = G\iint_s \frac{\rho}{\ell} ds \qquad [20]$$

In this case, V is a continuous function everywhere, but its first derivatives are discontinuous at the surface. Or, one can imagine two infinitesimally close density layers (double layer or layer of mass dipoles), where the units of density are now those of mass per area times length. It turns out that the potential in this case is given by (see Heiskanen and Moritz, 1967)

$$V_P = G\iint_s \rho\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) ds \qquad [21]$$

where $\partial/\partial n$ is the directional derivative along the perpendicular to the surface (Figure 5). Now, V itself is discontinuous at the surface, as are all its derivatives. In all cases, V is a Newtonian potential, being derived from the basic formula [11] for a point mass that follows from Newton's law of gravitation (eqn [6]).

The following properties of the gravitational potential are useful for subsequent expositions. First, consider Stokes' theorem, for a vector function, f, defined on a surface, s:

$$\iint_s (\nabla \times f)\cdot \boldsymbol{n}\, ds = \oint_p f \cdot dr \qquad [22]$$

where p is any closed path in the surface, $\boldsymbol{n}$ is the unit vector perpendicular to the surface, and $d\boldsymbol{r}$ is a differential displacement along the path. From eqn [9], we find

$$\nabla \times \boldsymbol{g} = 0 \qquad [23]$$

since $\boldsymbol{\nabla} \times \boldsymbol{\nabla} = \boldsymbol{0}$; and hence, applying Stokes' theorem, we find that with $\boldsymbol{F} = m\boldsymbol{g}$,

$$w = \oint F \cdot ds = 0 \qquad [24]$$

That is, the gravitational field is conservative: the work, w, expended in moving a mass around a closed path in this field vanishes. In contrast, dissipating forces (real forces!), like friction, expend work or energy, which shows again the special nature of the gravitational force.

It can be shown (Kellogg, 1953, p. 156) that the second partial derivatives of a Newtonian potential, V, satisfy the following differential equation, known as Poisson's equation:

$$\nabla^2 V = -4\pi G\rho \qquad [25]$$

where $\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}$ formally is the scalar product of two gradient operators and is called the Laplacian operator. In Cartesian coordinates, it is given by

$$\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \qquad [26]$$

Note that the Laplacian is a scalar operator. Equation [25] is a local characterization of the potential field, as opposed to the global characterization given by eqn [14]. Poisson's equation holds wherever the mass density, ρ, satisfies certain conditions similar to continuity (Hölder conditions, see Kellogg, 1953, pp. 152–153). A special case of eqn [25] applies for those points where the density vanishes (i.e., in free space); then, Poisson's equation turns into Laplace's equation:

$$\nabla^2 V = 0 \qquad [27]$$

It is easily verified that the point mass potential satisfies eqn [27]; that is,

$$\nabla^2 \left(\frac{1}{\ell}\right) = 0 \qquad [28]$$

where $\ell = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 - x_3')^2}$ and the mass point is at (x_1', x_2', x_3').

The solutions to Laplace's eqn [27] are known as harmonic functions; here, we also impose the conditions [10] on the solution, if it is a Newtonian potential and if the mass-free region includes infinity. Hence, every Newtonian potential is a harmonic function in free space. The converse is also true: every harmonic function can be represented as a Newtonian potential of a mass distribution (**Section 3.02.3.1**).

Whether as a volume or a layer density distribution, the corresponding potential is the sum or integral of the source value multiplied by the inverse distance function (or its normal derivative for the dipole layer). This function depends on both the source points and the computation point and is known as a Green's function. It is also known as the 'kernel' function of the integral representation of the potential. Functions of this type also play a dominant role in representing the potential as solutions to boundary-value problems (BVPs), as shown in subsequent sections.

This section concludes with another type of potential that finds uses in two-dimensional modeling. Many local geophysical density contrasts are long in one horizontal direction, such as faults and dikes, and may be characterized sufficiently in two dimensions, that is, with a single horizontal coordinate and the vertical coordinate. We start with the potential for an infinitesimally thin rod of finite length, $2c$, and parallel to the x_2-axis, whose standard Newtonian potential is

$$V = G\mu \int_{-c}^{c} \frac{dx_2'}{\sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 - x_3')^2}} \qquad [29]$$

where (x_1', x_3') is its offset from the origin (**Figure 6**) and μ is its constant line density. Taking the limit, $c \to \infty$, it can be shown that (up to a constant) the potential becomes (Kellogg, 1953, p. 56)

$$V = 2G\mu \ln \frac{1}{r} \qquad [30]$$

where $r = \sqrt{(x_1 - x_1')^2 + (x_3 - x_3')^2}$ is the distance of the computation point from the rod. This is sometimes referred to as the logarithmic potential. It is not a Newtonian potential since it does not vanish when $r \to \infty$ (see eqn [10]). If a long two-dimensional structure has a certain cross-sectional *shape* in the (x_1', x_3') plane, then its potential may be approximated by

$$V = 2G \iint_{\text{shape}} \mu \ln \frac{1}{r}\, dx_1' dx_3' \qquad [31]$$

and the derivatives follow easily for simple structures (see, e.g., Telford et al., 1990).

3.02.3 Boundary-Value Problems

If the density distribution of the Earth's interior and the boundary of the volume were known, then the problem of determining the Earth's gravitational potential field is solved by the volume integral of eqn [14]. In reality, of course, we do not have access to this information, at least not the density, with sufficient detail. The Preliminary Reference Earth Model, of Dziewonsky and Anderson (1981), still in use today by geophysicists, represents a good profile of Earth's radial density, but does not attempt to model in detail the lateral density heterogeneities. On the other hand, efforts to model the crust, especially at active tectonic plate margins where their dynamics is associated with earthquakes and mountain formation, often

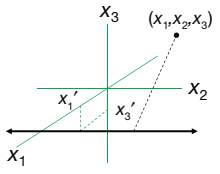


Figure 6 Geometry for the potential of a long rod.

rely (in addition to seismological data) on gravity data that are inverted to constrain the density distribution. Thus, gravity models, in principle, are developed not from three-dimensional density integrals, but from measurements on or above the Earth's surface. In this section, we see how the problem of determining the exterior gravitational potential can be reconstructed in terms of surface integrals, thus depending exclusively (at least in theory) on these measurements.

3.02.3.1 Green's Identities

Formally, eqn [27] represents a partial differential equation for V. Solving this equation is the essence of the determination of the Earth's external gravitational potential through potential theory. Like any differential equation, a complete solution is obtained only with the application of boundary conditions, that is, imposing values on the solution that it must assume at a boundary of the region in which it is valid. In our case, the boundary is the Earth's surface and the exterior space is where eqn [27] holds (the atmosphere and other celestial bodies are neglected or mathematically removed from the problem using corrections to the boundary values). In order to study the solutions to these BVPs, that is, to show that solutions exist and are unique, we take advantage of some very important theorems and identities. It is noted that only a rather elementary introduction to BVPs is offered here with no attempt to address the much larger field of solutions to partial differential equations.

The first, seminal result is Gauss' divergence theorem (analogous to Stokes' theorem, eqn [22]):

$$\iiint_v \nabla\cdot f \,\mathrm{d}v = \iint_s f_n \,\mathrm{d}s \qquad [32]$$

where f is an arbitrary (differentiable) vector function and $f_n = \boldsymbol{n}\cdot f$ is the component of f along the outward unit normal vector, $\boldsymbol{n}$ (see Figure 5). The surface, s, encloses the volume, v. Equation [32] says that the sum of how much f changes throughout the volume, that is, the net effect, ultimately, is equivalent to the sum of its values projected orthogonally with respect to the surface. Conceptually, a volume integral thus can be replaced by a surface integral, which is important since the gravitational potential is due to a volume density distribution that we do not know, but we do have access to gravitational quantities on a surface by way of measurements.

Equation [32] applies to general vector functions that have continuous first derivatives. In particular, let U and V be two scalar functions having continuous second derivatives; and consider the vector function $f = U\boldsymbol{\nabla}V$. Then, since $\boldsymbol{n}\cdot\boldsymbol{\nabla} = \partial/\partial n$, and

$$\boldsymbol{\nabla}\cdot(U\boldsymbol{\nabla}V) = \boldsymbol{\nabla}U\cdot\boldsymbol{\nabla}V + U\nabla^2 V \qquad [33]$$

we can apply Gauss' divergence theorem to get Green's first identity:

$$\iiint_v (\boldsymbol{\nabla}U\cdot\boldsymbol{\nabla}V + U\nabla^2 V)\,\mathrm{d}v = \iint_s U\frac{\partial V}{\partial n}\,\mathrm{d}s \qquad [34]$$

Interchanging the roles of U and V in eqn [34], one obtains a similar formula, which when subtracted from eqn [34] yields Green's second identity:

$$\iiint_v (U\nabla^2 V - V\nabla^2 U)\,\mathrm{d}v = \iint_s \left(U\frac{\partial V}{\partial n} - V\frac{\partial U}{\partial n}\right)\mathrm{d}s \qquad [35]$$

and which is valid for any U and V with continuous derivatives up to second order.

Now, let $U = 1/\ell$, where ℓ is the usual distance between an integration point and an evaluation point. And, suppose that the volume, v, is the space exterior to the Earth (i.e., Gauss' divergence theorem applies to any volume, not just volumes containing a mass distribution). With reference to Figure 7, consider the evaluation point, P, to be inside the volume (free space) that is bounded by the surface, s; P is thus outside the Earth's surface. Let V be a solution to eqn [27], that is, it is the gravitational potential of the Earth. From the volume, v, exclude the volume bounded by a small sphere, σ, centered at P. This sphere becomes part of the surface that bounds the volume, v. Then, since U, by our definition, is a point mass potential, $\nabla^2 U = 0$ everywhere in v (which excludes the interior of the small sphere around P); and the second identity [35] gives

$$\iint_s \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right)\mathrm{d}s + \iint_\sigma \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right)\mathrm{d}\sigma = 0 \qquad [36]$$

The unit vector, $\boldsymbol{n}$, represents the perpendicular pointing away from v. On the small sphere, $\boldsymbol{n}$ is opposite in direction to $\ell = r$, and the second integral becomes

$$\begin{aligned}\iint_\sigma \left(-\frac{1}{r}\frac{\partial V}{\partial r} + V\frac{\partial}{\partial r}\left(\frac{1}{r}\right)\right)\mathrm{d}\sigma &= -\iint_\Omega \frac{1}{r}\frac{\partial V}{\partial r}r^2\,\mathrm{d}\Omega - \iint_\Omega V\,\mathrm{d}\Omega \\ &= -\iint_\Omega \frac{\partial V}{\partial r}r\,\mathrm{d}\Omega - 4\pi\overline{V}\end{aligned} \qquad [37]$$

where $\mathrm{d}\sigma = r^2\mathrm{d}\Omega$; Ω is the solid angle, 4π; and $\overline{V}$ is an average value of V on σ. Now, in the limit as the radius of the small sphere shrinks to zero, the right-hand side of eqn [37] approaches $0 - 4\pi V_P$. Hence, eqn [36] becomes (Kellogg, 1953, p. 219)

$$V_P = \frac{1}{4\pi}\iint_s \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right)\mathrm{d}s \qquad [38]$$

with $\boldsymbol{n}$ pointing down (away from the space outside the Earth). This is a special case of Green's third identity. A change in sign of the right-side transforms $\boldsymbol{n}$ to a normal unit vector pointing into v, away from the masses, which conforms more to an Earth-centered coordinate system.

The right side of eqn [38] is the sum of single- and double-layer potentials and thus shows that every harmonic

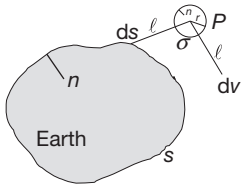


Figure 7 Geometry for special case of Green's third identity.

function (i.e., a function that satisfies Laplace's equation) can be written as a Newtonian potential. Equation [38] is also a solution to a BVP; in this case, the boundary values are independent values of V and of its normal derivative, both on s (Cauchy problem). In the succeeding text and in Section 3.02.4.1, we encounter another BVP (Robin problem) in which the potential and its normal derivative are given in a specified linear combination on s. Using a similar procedure and with some extra care, it can be shown (see also Courant and Hilbert, 1962, vol. 2, p. 256 (footnote)) that if P is on the surface, then

$$V_P = \frac{1}{2\pi}\iint_s \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right)\right) ds \qquad [39]$$

where $\boldsymbol{n}$ points into the masses. Comparing this to eqn [38], we see that V is discontinuous as one approaches the surface; this is due to the double-layer part (see eqn [21]).

Equation [38] demonstrates that a solution to a particular BVP exists. Specifically, we are able to measure the potential (up to a constant) and its derivatives (the gravitational acceleration) on the surface and thus have a formula to compute the potential anywhere in exterior space, provided that we also know the surface, s. Other BVPs also have solutions under appropriate conditions; a discussion of existence theorems is beyond the present scope and may be found in Kellogg (1953, Chapter 1.13). Equation [39] has deep geodetic significance. One objective of geodesy is to determine the shape of the Earth's surface. If we have measurements of gravitational quantities on the Earth's surface, then conceptually we are able to determine its shape from eqn [39], where it would be the only unknown quantity. This is the basis behind the work of Molodensky et al. (1962) to which we return briefly at the end of this section.

3.02.3.2 Uniqueness Theorems

Often the existence of a solution is proved simply by finding one (as illustrated earlier). Whether such as solution is the only one depends on a corresponding uniqueness theorem. That is, we wish to know if a certain set of boundary values will yield just one potential in space. Before considering such theorems, we classify the BVPs that are typically encountered when determining the exterior potential from measurements on a boundary. In all cases, it is an exterior BVP; that is, the gravitational potential, V, is harmonic ($\nabla^2 V = 0$) in the space exterior to a closed surface that contains all the masses. The exterior space thus contains infinity. Interior BVPs can be constructed, as well, but are not applicable to our objectives:

- Dirichlet problem (or BVP of the first kind): solve for V in the exterior space, given its values everywhere on the boundary.
- Neumann problem (or BVP of the second kind): solve for V in the exterior space, given values of its normal derivative everywhere on the boundary.
- Robin problem (mixed BVP or BVP of the third kind): solve for V in the exterior space, given a linear combination of it and its normal derivative on the boundary.

Using Green's identities, we prove the following theorems for these exterior problems; similar results hold for the interior problems.

Theorem 1 *If V is harmonic (hence continuously differentiable) in a closed region, v, and if V vanishes everywhere on the boundary, s, then V also vanishes everywhere in the region, v.*

Proof: Since $V = 0$ on s, Green's first identity (eqn [34]) with $U = V$ gives

$$\iiint_v (\boldsymbol{\nabla} V)^2 dv = \iint_s V\frac{\partial V}{\partial n} ds = 0 \qquad [40]$$

where the integral on the left side is, therefore, always zero; and the integrand is always nonnegative. Hence, it must vanish, $\nabla V = 0$, everywhere in v, which implies that $V =$ constant in v. Since V is continuous in v and $V = 0$ on s, that constant must be zero; and so, $V = 0$ in v.

This theorem solves the Dirichlet problem for the trivial case of zero boundary values, and it enables the following uniqueness theorem for the general Dirichlet problem.

Theorem 2 (Stokes' theorem) *If V is harmonic (hence continuously differentiable) in a closed region, v, then V is uniquely determined in v by its values on the boundary, s.*

Proof: Suppose the determination is not unique: that is, suppose there are V_1 and V_2 both harmonic in v and both having the same boundary values on s. Then, the function $V = V_2 - V_1$ is harmonic in v with all boundary values equal to zero. Hence, by Theorem 1, $V_2 - V_1 = 0$ identically in v, or $V_2 = V_1$ everywhere, which implies that any determination is unique based on the boundary values.

Theorem 3 *If V is harmonic (hence continuously differentiable) in the exterior region, v, with closed boundary, s, then V is uniquely determined by the values of its normal derivative on s.*

Proof: We begin with Green's first identity (eqn [34]) as in the proof of Theorem 1 to show that if the normal derivative vanishes everywhere on s, then V is a constant in v. Now, suppose there are two harmonic functions in v, V_1 and V_2, with the same normal derivative values on s. Then, the normal derivative values of their difference are zero; and by the aforementioned demonstration, $V = V_2 - V_1 =$ constant in v. Since V is a Newtonian potential in the exterior space, that constant is zero, since by eqn [10], $\lim_{\ell\to\infty} V = 0$. Thus, $V_2 = V_1$, and the boundary values determine the potential uniquely.

This is a uniqueness theorem for the exterior Neumann BVP. Solutions to the interior problem are unique only up to an arbitrary constant.

Theorem 4 *Suppose V is harmonic (hence continuously differentiable) in the closed region, v, with boundary, s; and suppose the boundary values are given by*

$$g = \alpha V|_s + \beta\left.\frac{\partial V}{\partial n}\right|_s \qquad [41]$$

then, V is uniquely determined by these values if $\alpha/\beta > 0$.

Proof: Suppose there are two harmonic functions, V_1 and V_2, with the same boundary values, g, on s. Then, $V = V_2 - V_1$ is harmonic with boundary values

$$\alpha(V_2 - V_1)|_s + \beta\left(\frac{\partial V_2}{\partial n} - \frac{\partial V_1}{\partial n}\right)\Big|_s = 0 \qquad [42]$$

where, with $U = V = V_2 - V_1$, Green's first identity (eqn [34]) gives

$$\iiint_v (\boldsymbol{\nabla}(V_2 - V_1))^2 \mathrm{d}v = \iint_s (V_2 - V_1)\frac{-\alpha}{\beta}(V_2 - V_1)\mathrm{d}s \qquad [43]$$

Then,

$$\iiint_v (\boldsymbol{\nabla}(V_2 - V_1))^2 \mathrm{d}v + \frac{\alpha}{\beta}\iint_S (V_2 - V_1)^2 \mathrm{d}S = 0 \qquad [44]$$

and since $\alpha/\beta > 0$, eqn [44] implies that $\boldsymbol{\nabla}(V_2 - V_1) = 0$ in v; and $V_2 - V_1 = 0$ on s. Hence, $V_2 - V_1 =$ constant in v; and $V_2 = V_1$ on s. By the continuity of V_1 and V_2, the constant must be zero; and the uniqueness is proved.

The solution to the Robin problem is thus unique only in certain cases. The most famous problem in physical geodesy is the determination of the disturbing potential, T, from gravity anomalies, Δg, on the geoid (**Section 3.02.1.3**). Suppose T is harmonic outside the geoid; the second of eqn [5] provides an approximate form of boundary condition, showing that this is a type of Robin problem. We find that $\alpha = -2/r$; and recalling that when v is the exterior space, the unit vector $\boldsymbol{n}$ points inward toward the masses, that is, $\partial/\partial n = -\partial/\partial r$, we get $\beta = 1$. Thus, the condition in Theorem 4 on α/β is not fulfilled and the uniqueness is not guaranteed. In fact, we will see that the solution obtained for the spherical boundary is arbitrary with respect to the coordinate origin (**Section 3.02.4.1**).

3.02.3.3 Solutions by Integral Equation

Green's identities show how a solution to Laplace's equation can be transformed from a volume integral, that is, an integral of source points, to a surface integral of boundary-value points, as demonstrated by eqn [38]. The uniqueness theorems for the BVPs suggest that the potential due to a volume density distribution can also be represented as due to a generalized density layer on the bounding surface, as long as the result is harmonic in exterior space, satisfies the boundary conditions, and is regular at infinity like a Newtonian potential. Molodensky et al. (1962) supposed that the disturbing potential is expressible as

$$T = \iint_s \frac{\mu}{\ell}\mathrm{d}s \qquad [45]$$

where μ is a surface density to be solved using the boundary condition. With the spherical approximation for the gravity anomaly (eqn [5]), one arrives at the following integral equation:

$$2\pi\mu\cos\zeta - \iint_s \left(\frac{\partial}{\partial r}\left(\frac{1}{\ell}\right) + \frac{2}{r\ell}\right)\mu\,\mathrm{d}s = \Delta g \qquad [46]$$

The first term accounts for the discontinuity at the surface of the derivative of the potential of a density layer, where ζ is the deflection angle between the normal to the surface and the direction of the (radial) derivative (Günter, 1967, p. 69; Heiskanen and Moritz, 1967, p. 6). This Fredholm integral equation of the second kind can be simplified with further approximations; and a solution for the density, μ, ultimately leads to the solution for the disturbing potential (Moritz, 1980).

Other forms of representing the potential have also been investigated, where Green's functions other than $1/\ell$ lead to simplifications of the integral equation (e.g., Petrovskaya, 1979). Nevertheless, most practical solutions rely on approximations, such as the spherical approximation for the boundary condition, and even the formulated solutions are not strictly guaranteed to converge to the true solutions (Moritz, 1980). Further treatments of the BVP in a geodetic/mathematical setting may be found in the volume edited by Sansò and Rummel (1997).

In the next section, we consider solutions for T as surface integrals of boundary values with appropriate Green's functions and spherical boundaries. In other words, the boundary values (whether of the first, second, or third kind) may be thought of as sources and the consequent potential is again the sum (integral) of the product of a boundary value and an appropriate Green's function (i.e., a function that depends on both the source point and the computation point in some form of inverse distance in accordance with Newtonian potential theory). Solutions are readily obtained because the boundary is a sphere.

3.02.4 Solutions to the Spherical BVP

This section develops two types of solutions to standard BVPs when the boundary is a sphere: the spherical harmonic series and an integral with a Green's function. All three types of problems are solved, but emphasis is put on the third BVP since gravity anomalies are the most prevalent boundary values (on land, at least). In addition, it is shown how Green's function integrals can be inverted to obtain, for example, gravity anomalies from values of the potential, now considered as boundary values. Not all possible inverse relationships are given, but it should be clear at the end that, in principle, virtually any gravitational quantity can be obtained in space from any other quantity on the spherical boundary.

3.02.4.1 Spherical Harmonics and Green's Functions

For simple boundaries, Laplace's equation [27] is relatively easy to solve provided that there is an appropriate coordinate system. For the Earth, the solutions commonly rely on approximating the boundary by a sphere. This case is described in detail and a more accurate approximation based on an ellipsoid of revolution is briefly examined in **Section 3.02.5.2** for the normal potential. In spherical polar coordinates, (θ,λ,r), the Laplacian operator is given by (Hobson, 1965, p. 9)

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\lambda^2} \qquad [47]$$

A solution to $\nabla^2 V = 0$ in the space outside a sphere of radius, R, with center at the coordinate origin can be found by the method of separation of variables, whereby one postulates the form of the solution, V, as

$$V(\theta,\lambda,r) = f(\theta)g(\lambda)h(r) \qquad [48]$$

Substituting this and the Laplacian, eqn [47], into eqn [27], the multivariate partial differential equation separates into three univariate ordinary differential equations (Hobson, 1965, p. 9; Morse and Feshbach, 1953, p. 1264). Their solutions are well-known functions, for example,

$$V(\theta,\lambda,r) = P_{nm}(\cos\theta)\sin m\lambda\frac{1}{r^{n+1}} \quad \text{or}$$
$$V(\theta,\lambda,r) = P_{nm}(\cos\theta)\cos m\lambda\frac{1}{r^{n+1}} \qquad [49]$$

where $P_{nm}(t)$ is the associated Legendre function of the first kind and n and m are integers such that $0 \le m \le n$ and $n \ge 0$. Other solutions are also possible (e.g., $g(\lambda) = e^{a\lambda}$ $(a \varepsilon \mathbb{R})$ and $h(r) = r^n$), but only eqn [49] is consistent with the problem at hand: to find a real-valued Newtonian potential for the exterior space of the Earth (regular at infinity and 2π-periodic in longitude). The general solution is a linear combination of solutions of the forms given by eqn [49] for all possible integers, n and m, and can be written compactly as

$$V(\theta,\lambda,r) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+1} v_{nm}\overline{Y}_{nm}(\theta,\lambda) \qquad [50]$$

where the $\overline{Y}_{nm}$ are surface spherical harmonic functions defined as

$$\overline{Y}_{nm}(\theta,\lambda) = \overline{P}_{n|m|}(\cos\theta)\begin{cases}\cos m\lambda, & m \ge 0\\ \sin|m|\lambda, & m < 0\end{cases} \qquad [51]$$

and $\overline{P}_{nm}$ is a normalization of P_{nm} so that the orthogonality of the spherical harmonics is simply

$$\frac{1}{4\pi}\iint_{\sigma}\overline{Y}_{nm}(\theta,\lambda)\overline{Y}_{n'm'}(\theta,\lambda)\mathrm{d}\sigma = \begin{cases}1, & n = n' \text{ and } m = m'\\ 0, & n \ne n' \text{ or } m \ne m'\end{cases} \qquad [52]$$

and where $\sigma = \{(\theta,\lambda)|0 \le \theta \le \pi,\ 0 \le \lambda \le 2\pi\}$ represents the unit sphere, with $\mathrm{d}\sigma = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\lambda$. For a complete mathematical treatment of spherical harmonics, one may refer to Müller (1966). The bounding spherical radius, R, is introduced so that all the constant coefficients, v_{nm}, sometimes known as Stokes' constants (**Section 3.02.5.1**), have identical units of measure. Applying the orthogonality to the general solution [50], these coefficients can be determined if the function, V, is known on the bounding sphere (boundary condition):

$$v_{nm} = \frac{1}{4\pi}\iint_{\sigma}V(\theta,\lambda,R)\overline{Y}_{nm}(\theta,\lambda)\mathrm{d}\sigma \qquad [53]$$

Equation [50] is known as a spherical harmonic expansion of V, and with eqn [53], it represents a solution to the Dirichlet BVP if the boundary is a sphere. The solution thus exists and is unique in the sense that these boundary values generate no other potential. We will, however, find another equivalent form of the solution.

In a more formal mathematical setting, the solution [52] is an infinite linear combination of orthogonal basis functions (eigenfunctions), and the coefficients, v_{nm}, are the corresponding eigenvalues. One may also interpret the set of coefficients as the spectrum (Legendre spectrum) of the potential on the sphere of radius, R (analogous to the Fourier spectrum of a function on the plane or line). The integers, n and m, correspond to wave numbers and are called degree (n) and order (m), respectively. The spherical harmonics are further classified as zonal ($m = 0$), meaning that the zeros of $\overline{Y}_{n0}$ divide the sphere into latitudinal zones; sectorial ($m = n$), where the zeros of $\overline{Y}_{nn}$ divide the sphere into longitudinal sectors; and tesseral (the zeros of $\overline{Y}_{nm}$ tessellate the sphere) (**Figure 8**).

While the spherical harmonic series has its advantages in global representations and spectral interpretations of the field, a Green's function representation provides a more local characterization of the field. That is, changing a boundary value anywhere on the globe changes all coefficients, v_{nm}, according to eqn [53], which poses a numerical challenge both in applications and in keeping a standard model up-to-date. However, since Green's function essentially depends on the inverse distance (or higher powers thereof), a remote change in boundary value generally does not appreciably affect the local determination of the field.

When the boundary is a sphere, the solutions to the BVPs using a Green's function are easily derived from the spherical harmonic series representation. Moreover, it is possible to derive additional integral relationships (with appropriate Green's functions) among all the derivatives of the potential. To formalize and simultaneously simplify these derivations, consider harmonic functions, f and h, where h depends only

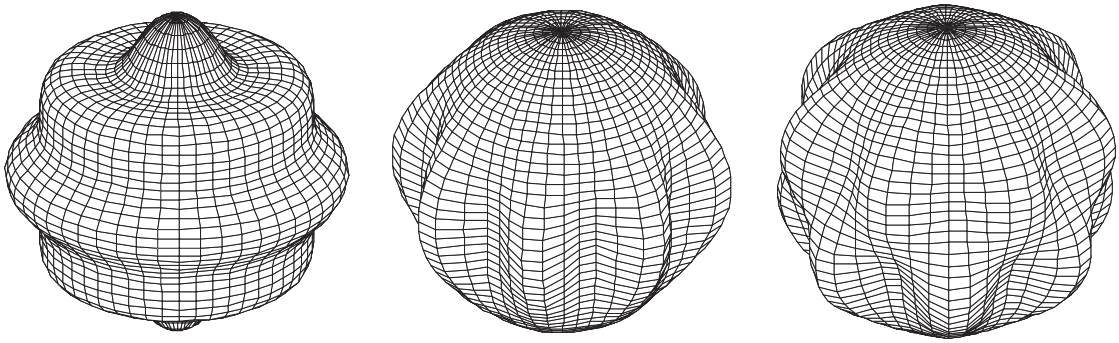


Figure 8 Examples of zonal, sectorial, and tesseral harmonics on the sphere.

on θ and r, and the function g, defined on the sphere of radius, R. Thus, let

$$f(\theta,\lambda,r)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+1}f_{nm}\overline{Y}_{nm}(\theta,\lambda),\quad r\geq R \qquad [54]$$

$$h(\theta,r)=\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1}h_nP_n(\cos\theta),\quad r\geq R \qquad [55]$$

$$g(\theta,\lambda,R)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}g_{nm}\overline{Y}_{nm}(\theta,\lambda) \qquad [56]$$

where $P_n(\cos\theta)=\overline{P}_{n0}(\cos\theta)/\sqrt{2n+1}$ is the nth-degree Legendre polynomial. Constants f_{nm} and g_{nm} are the respective harmonic coefficients of f and g when these functions are restricted to the sphere of radius, R. Then, using the decomposition formula for Legendre polynomials,

$$P_n(\cos\psi)=\frac{1}{2n+1}\sum_{m=-n}^{n}\overline{Y}_{nm}(\theta,\lambda)\overline{Y}_{nm}\left(\theta',\lambda'\right) \qquad [57]$$

where

$$\cos\psi=\cos\theta\cos\theta'+\sin\theta\sin\theta'\cos\left(\lambda-\lambda'\right) \qquad [58]$$

it is easy to prove the following theorem.

Theorem 5 (convolution theorem in spectral analysis on the sphere)

$$f(\theta,\lambda,r)=\frac{1}{4\pi}\iint_{\sigma}g\left(\theta',\lambda',R\right)h(\psi,r)\mathrm{d}\sigma \quad \text{if and only if}$$

$$f_{nm}=g_{nm}h_n \qquad [59]$$

where here and in the following, $\mathrm{d}\sigma=\sin\theta'\,\mathrm{d}\theta'\,\mathrm{d}\lambda'$. The angle, ψ, is the distance on the unit sphere between points (θ,λ) and (θ',λ').

Proof: The forward statement [59] follows directly by substituting eqns [55] and [57] into the first eqn [59], together with the spherical harmonic expansion [56] for g. A comparison with the spherical harmonic expansion for f yields the result. All steps in this proof are reversible, and so, the reverse statement also holds.

Consider now f to be the potential, V, on and outside the sphere of radius, R, and its restriction to the sphere to be the function, $g(\theta,\lambda)=V(\theta,\lambda,R)$. Then, since $g_{nm}=f_{nm}$, we have $h_n=1$, for all n; and by Theorem 5

$$V(\theta,\lambda,r)=\frac{1}{4\pi}\iint_{\sigma}V\left(\theta',\lambda',R\right)U(\psi,r)\mathrm{d}\sigma \qquad [60]$$

where

$$U(\psi,r)=\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi) \qquad [61]$$

For the distance,

$$\ell=\sqrt{r^2+R^2-2rR\cos\psi} \qquad [62]$$

between points (θ,λ,r) and (θ',λ',R), with $r\geq R$, the identity (the *Coulomb expansion*; Cushing, 1975, p. 155)

$$\frac{1}{\ell}=\frac{1}{R}\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi) \qquad [63]$$

yields after some arithmetic (based on taking the derivative on both sides with respect to r)

$$U(\psi,r)=\frac{R(r^2-R^2)}{\ell^3} \qquad [64]$$

Solutions [50] and [60] to the Dirichlet BVP for a spherical boundary are identical in view of the convolution theorem [59]. The integral in [60] is known as *Poisson's integral*, and the function U is the corresponding Green's function, also known as Poisson's kernel.

For convenience, one separates Earth's gravitational potential into a reference potential (**Section 3.02.1.3**) and the disturbing potential, T. The disturbing potential is harmonic in free space and satisfies Poisson's integral if the boundary is a sphere. In deference to physical geodesy where relationships between the disturbing potential and its derivatives are routinely applied, the following derivations are developed in terms of T but hold equally for any exterior Newtonian potential. Let

$$T(\theta,\lambda,r)=\frac{GM}{R}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+1}\delta C_{nm}\overline{Y}_{nm}(\theta,\lambda) \qquad [65]$$

where M is the total mass (including the atmosphere) of the Earth and the δC_{nm} are unit-less harmonic coefficients, being also the differences between coefficients for the total and reference gravitational potentials (**Section 3.02.5.2**). The coefficient, δC_{00}, is zero under the assumption that the reference field accounts completely for the central part of the total field. Note also that these coefficients specifically refer to the sphere of radius, R, that is, they constitute the (normalized) spectrum of T on this sphere.

The gravity disturbance is defined in spherical approximation to be the negative radial derivative of T, the first of eqn [5]. From eqn [65], we have

$$\begin{aligned}\delta g(\theta,\lambda,r)&=-\frac{\partial}{\partial r}T(\theta,\lambda,r)\\&=\frac{GM}{R^2}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+2}(n+1)\delta C_{nm}\overline{Y}_{nm}(\theta,\lambda)\end{aligned} \qquad [66]$$

and applying the convolution theorem [59] to $f(\theta,\lambda,r)=T(\theta,\lambda,r)$ and $g(\theta,\lambda,R)=R\delta g(\theta,\lambda,R)$, we obtain

$$T(\theta,\lambda,r)=\frac{R}{4\pi}\iint_{\sigma}\delta g\left(\theta',\lambda',R\right)H(\psi,r)\mathrm{d}\sigma \qquad [67]$$

where with $g_{nm}=GM(n+1)\delta C_{nm}/R$ and $f_{nm}=GM\delta C_{nm}$, we have $h_n=f_{nm}/g_{nm}=R/(n+1)$ and hence

$$H(\psi,r)=\sum_{n=0}^{\infty}\frac{2n+1}{n+1}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi) \qquad [68]$$

The integral in [67] is known as the *Hotine integral*, Green's function, H, is called the Hotine kernel, and with a derivation based on eqn [63], it is given by (Hotine, 1969, p. 311)

$$H(\psi,r)=\frac{2R}{\ell}-\ln\left(\frac{R+\ell-r\cos\psi}{r(1-\cos\psi)}\right) \qquad [69]$$

Equation [67] solves the Neumann BVP when the boundary is a sphere.

The gravity anomaly, again, in spherical approximation is defined by (eqn [5])

$$\Delta g(\theta,\lambda,r) = \left(-\frac{\partial}{\partial r} - \frac{2}{r}\right) T(\theta,\lambda,r) \qquad [70]$$

or also

$$\Delta g(\theta,\lambda,r) = \frac{GM}{R^2}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+2}(n-1)\delta C_{nm}\overline{Y}_{nm}(\theta,\lambda) \qquad [71]$$

where in this case, we have $g_{nm} = GM(n-1)\delta C_{nm}/R$ and $h_n = R/(n-1)$. The convolution theorem then leads to the geodetically famous *Stokes' integral*:

$$T(\theta,\lambda,r) = \frac{R}{4\pi}\iint_{\sigma}\Delta g\left(\theta',\lambda',R\right)S(\psi,r)\mathrm{d}\sigma \qquad [72]$$

where we *define* the corresponding Green's function

$$S(\psi,r) = \sum_{n=2}^{\infty}\frac{2n+1}{n-1}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi) = 2\frac{R}{\ell} + \frac{R}{r} - 3\frac{R\ell}{r^2} - 5\frac{R^2}{r^2}\cos\psi - 3\frac{R^2}{r^2}\cos\psi\ln\frac{\ell + r - R\cos\psi}{2r} \qquad [73]$$

more commonly called Stokes' kernel. Equation [72] solves the Robin BVP if the boundary is a sphere, but it includes specific constraints that ensure the solution's uniqueness – the solution by itself is not unique, in this case, as proved in Section 3.02.3.2. Indeed, eqn [71] shows that the gravity anomaly has no first-degree harmonics that for the disturbing potential, therefore, cannot be determined from the boundary values. Conventionally, Stokes' kernel also excludes the zero-degree harmonic, and thus, the complete solution for the disturbing potential is given by

$$T(\theta,\lambda,r) = \frac{GM}{r}\delta C_{00} + \frac{GM}{R}\sum_{m=-1}^{1}\left(\frac{R}{r}\right)^{2}\delta C_{1m}\overline{Y}_{1m}(\theta,\lambda) + \frac{R}{4\pi}\iint_{\sigma}\Delta g\left(\theta',\lambda',R\right)S(\psi,r)\mathrm{d}\sigma \qquad [74]$$

The central term, δC_{00}, is proportional to the difference in GM of the Earth and of the reference ellipsoid and is zero to high accuracy. The first-degree harmonic coefficients, δC_{1m}, are proportional to the center of mass coordinates and can also be set to zero with appropriate definition of the coordinate system (see Section 3.02.5.1). Thus, Stokes' integral [72] is the more common expression for the disturbing potential, but it embodies hidden constraints.

We note that gravity anomalies can also serve as boundary values in the harmonic series form of the solution for the disturbing potential. Applying the orthogonality of the spherical harmonics to eqn [71] yields immediately

$$\delta C_{nm} = \frac{R^2}{4\pi(n-1)GM}\iint_{\sigma}\Delta g\left(\theta',\lambda',R\right)\overline{Y}_{nm}\left(\theta',\lambda'\right)\mathrm{d}\sigma, \quad n \geq 2 \qquad [75]$$

A similar formula holds when gravity disturbances are the boundary values ($n-1$ in the denominator changes to $n+1$). In either case, the boundary values formally are assumed to reside on a sphere of radius, R. An approximation results if they are given on the geoid, as is usually the case.

The two solutions to the Robin BVP (Section 3.02.3.2) with boundary values defined by eqn [70], one given by eqn [72] and the other by eqns [65] and [75], are equivalent. Similar to Poisson's integral, eqn [72] may be viewed more as a local representation since Stokes' function, $S(\psi,r)$, attenuates significantly with distance from the computation point. Equation [65], on the other hand, is a global representation, where the harmonic coefficients, δC_{nm}, may also be determined from other boundary values, for example, *in situ* satellite measurements or from satellite tracking data (Section 3.02.6.1). In practice, the two solutions are combined in order to reduce the error when necessarily limiting the integral of the latter to a local neighborhood. One removes a global model (finite series) from the gravity anomalies and subsequently restores the model to the disturbing potential. This remove–restore technique is discussed in Section 3.02.6.2.

3.02.4.2 Inverse Stokes' and Hotine Integrals

The convolution integrals in the previous section can easily be inverted by considering again the spectral relationships. For the gravity anomaly in space, we note that $f = r\Delta g$ is harmonic with coefficients, $f_{nm} = GM(n-1)\delta C_{nm}/R$. Letting $g = T|_{r=R}$ with $g_{nm} = GM\delta C_{nm}/R$, we find that $h_n = n-1$; and by the convolution theorem, we can write

$$\Delta g(\theta,\lambda,r) = \frac{1}{4\pi R}\iint_{\sigma}T\left(\theta',\lambda',R\right)\hat{Z}(\psi,r)\mathrm{d}\sigma \qquad [76]$$

where

$$\hat{Z}(\psi,r) = \sum_{n=0}^{\infty}(2n+1)(n-1)\left(\frac{R}{r}\right)^{n+2}P_n(\cos\psi) \qquad [77]$$

The zero- and first-degree terms are included provisionally. Note that

$$\hat{Z}(\psi,r) = -R\left(\frac{\partial}{\partial r} + \frac{2}{r}\right)U(\psi,r) \qquad [78]$$

That is, we could have simply used the Dirichlet solution [60] to obtain the gravity anomaly, as given by [76], from the disturbing potential. It is convenient to separate the kernel function as follows:

$$\hat{Z}(\psi,r) = Z(\psi,r) - \sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+2}P_n(\cos\psi) \qquad [79]$$

where

$$Z(\psi,r) = \sum_{n=1}^{\infty}(2n+1)n\left(\frac{R}{r}\right)^{n+2}P_n(\cos\psi) \qquad [80]$$

and we find that with eqn [60] applied to T,

$$\Delta g(\theta,\lambda,r) = -\frac{1}{r}T(\theta,\lambda,r) + \frac{1}{4\pi R}\iint_{\sigma}T\left(\theta',\lambda',R\right)Z(\psi,r)\,\mathrm{d}\sigma \qquad [81]$$

Now, since Z has no zero-degree harmonics, its integral over the sphere vanishes, and one can write the numerically more convenient formula:

$$\Delta g(\theta,\lambda,r) = -\frac{1}{r}T(\theta,\lambda,r) + \frac{1}{4\pi R}\iint_{\sigma}\Big(T\left(\theta',\lambda',R\right) - T(\theta,\lambda,R)\Big)Z(\psi,r)\,\mathrm{d}\sigma \quad [82]$$

This is the inverse Stokes' formula. Given T on the sphere of radius R, for example, in the form of geoid undulations, $T=\gamma N$, this form is useful when the gravity anomaly is also desired on this sphere. It is one way to determine gravity anomalies on the ocean surface from satellite altimetry, which yields N directly, given the position of the satellite. Analogously, from eqns [66] and [70], it is readily seen that the inverse Hotine formula is given by

$$\delta g(\theta,\lambda,r) = \frac{1}{r}T(\theta,\lambda,r) + \frac{1}{4\pi R}\iint_{\sigma}\Big(T\left(\theta',\lambda',R\right) - T(\theta,\lambda,R)\Big)Z(\psi,r)\,\mathrm{d}\sigma \quad [83]$$

Note that the sum of eqns [82] and [83] yields the approximate relationship between the gravity disturbance and the gravity anomaly inferred from eqn [5]. A closed expression for Z may be obtained by writing $(2n+1)n=(n+1)(n+2)+n(n+1)-3(n+1)+1$ and combining appropriate derivatives of eqn [63]:

$$Z(\psi,r) = \frac{R^2}{\ell^5}\left(\frac{\ell^4}{r} + \ell^2(5R\cos\psi - r) - 6rR^2\sin^2\psi\right) \quad [84]$$

Finally, we realize that, for $r=R$, the series for $Z(\psi,R)$ is not uniformly convergent and special numerical procedures (that are outside the present scope) are required to approximate the corresponding integrals.

3.02.4.3 Vening-Meinesz Integral and Its Inverse

Other derivatives of the disturbing potential may also be determined from boundary values. We consider here only gravity anomalies, being the most prevalent data type on land areas. The solution is in the form of either a series – simply the derivative of the series [65] with coefficients given by eqn [75] – or an integral with appropriate derivative of Green's function. The horizontal derivatives of the disturbing potential may be interpreted as the deflections of the vertical to which they are proportional in spherical approximation (eqn [5]):

$$\begin{Bmatrix}\xi(\theta,\lambda,r)\\ \eta(\theta,\lambda,r)\end{Bmatrix} = \frac{1}{\gamma(\theta,r)}\begin{Bmatrix}\frac{1}{r}\frac{\partial}{\partial\theta}\\ -\frac{1}{r\sin\theta}\frac{\partial}{\partial\lambda}\end{Bmatrix}T(\theta,\lambda,r) \quad [85]$$

where ξ,η are the north and east deflection components, respectively, and γ is the normal gravity. Clearly, the derivatives can be taken directly inside Stokes' integral, and we find

$$\begin{Bmatrix}\xi(\theta,\lambda,r)\\ \eta(\theta,\lambda,r)\end{Bmatrix} = \frac{R}{4\pi r\gamma(\theta,r)}\iint_{\sigma}\Delta g\left(\theta',\lambda',R\right)\frac{\partial}{\partial\psi}S(\psi,r)\begin{Bmatrix}\cos\alpha\\ \sin\alpha\end{Bmatrix}\mathrm{d}\sigma \quad [86]$$

where

$$\begin{Bmatrix}\frac{1}{r}\frac{\partial}{\partial\theta}\\ -\frac{1}{r\sin\theta}\frac{\partial}{\partial\lambda}\end{Bmatrix} = \frac{1}{r}\begin{Bmatrix}\frac{\partial\psi}{\partial\theta}\\ -\frac{1}{\sin\theta}\frac{\partial\psi}{\partial\lambda}\end{Bmatrix}\frac{\partial}{\partial\psi} \quad [87]$$

and

$$\frac{\partial\psi}{\partial\theta} = \frac{1}{\sin\psi}\left(\sin\theta\cos\theta' - \cos\theta\sin\theta'\cos\left(\lambda-\lambda'\right)\right) = \cos\alpha$$
$$-\frac{1}{\sin\theta}\frac{\partial\psi}{\partial\lambda} = \frac{1}{\sin\psi}\sin\theta'\sin\left(\lambda'-\lambda\right) = \sin\alpha \quad [88]$$

where the angle, α, is the azimuth of (θ',λ') at (θ,λ) on the unit sphere. The integrals [86] are known as the *Vening-Meinesz integrals*. Analogous integrals for the deflections arise when the boundary values are the gravity disturbances (Green's functions are then derivatives of the Hotine kernel).

For the inverse Vening-Meinesz integrals, we need to make use of Green's first identity for surface functions, f and g:

$$\iint_{s} f\Delta^*(g)\,\mathrm{d}s + \iint_{s}\boldsymbol{\nabla}f\cdot\boldsymbol{\nabla}g\,\mathrm{d}s = \int_{b} f\boldsymbol{\nabla}g\cdot\boldsymbol{n}\,\mathrm{d}b \quad [89]$$

where b is the boundary (a line) of a surface, s; $\boldsymbol{\nabla}$ and Δ^* are the gradient and Laplace–Beltrami operators, which for the sphere are given by

$$\boldsymbol{\nabla} = \begin{pmatrix}\frac{\partial}{\partial\theta}\\ \frac{1}{\sin\theta}\frac{\partial}{\partial\lambda}\end{pmatrix}, \quad \Delta^* = \frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\lambda^2} \quad [90]$$

and $\boldsymbol{n}$ is the unit vector normal to b. For a closed surface such as the sphere, the line integral vanishes, and we have

$$\iint_{\sigma} f\Delta^*(g)\mathrm{d}\sigma = -\iint_{\sigma}\boldsymbol{\nabla}f\cdot\boldsymbol{\nabla}g\;\mathrm{d}\sigma \quad [91]$$

The surface spherical harmonics, $\overline{Y}_{nm}(\theta,\lambda)$, satisfy the following differential equation:

$$\Delta^*\overline{Y}_{nm}(\theta,\lambda) + n(n+1)\overline{Y}_{nm}(\theta,\lambda) = 0 \quad [92]$$

and, therefore, the harmonic coefficients of $-\Delta^*T(\theta,\lambda,r)$ on the sphere of radius, R, are

$$g_{nm} \equiv \left[-\Delta^*T(\theta,\lambda,r)\right]_{nm} = n(n+1)\frac{GM}{R}\delta C_{nm} \quad [93]$$

Hence, by the convolution theorem (again, considering the harmonic function, $f=r\Delta g$),

$$\Delta g(\theta,\lambda,r) = -\frac{1}{4\pi R}\iint_{\sigma}\Delta^*T\left(\theta',\lambda',R\right)W(\psi,r)\mathrm{d}\sigma \quad [94]$$

where, since $f_{nm}=GM(n-1)\delta C_{nm}/R$,

$$W(\psi,r) = \sum_{n=2}^{\infty}\frac{(2n+1)(n-1)}{n(n+1)}\left(\frac{R}{r}\right)^{n+2}P_n(\cos\psi) \quad [95]$$

and where the zero-degree term of the gravity anomaly must be treated separately (e.g., it is set to zero in this case). Using Green's identity [91] and eqns [85], [87], and [88], we have

$$\Delta g(\theta,\lambda,r) = \frac{1}{4\pi R}\iint_{\sigma} \nabla T\left(\theta',\lambda',R\right)\cdot\nabla W(\psi,r)\mathrm{d}\sigma$$
$$= \frac{\gamma_0}{4\pi}\iint_{\sigma}\left(\xi\left(\theta',\lambda',R\right)\cos\alpha+\eta\left(\theta',\lambda',R\right)\sin\alpha\right)$$
$$\frac{\partial}{\partial\psi}W(\psi,r)\mathrm{d}\sigma \qquad [96]$$

where normal gravity on the sphere of radius, R, is approximated as a constant: $\gamma(\theta,R)\simeq\gamma_0$. Equation [96] represents a second way to compute gravity anomalies from satellite altimetry, where the along-track and cross-track altimetric differences are used to approximate the deflection components (with appropriate rotation to north and east components). Employing differences in altimetric measurements benefits the estimation since systematic errors, such as orbit error, cancel out. To speed up the computations, the problem is reformulated in the spectral domain; see, for example, Sandwell and Smith (1996).

Clearly, following the same procedure for $f=T$, we also have the following relationship:

$$T(\theta,\lambda,r) = \frac{R\gamma_0}{4\pi}\iint_{\sigma}\left(\xi\left(\theta',\lambda',R\right)\cos\alpha+\eta\left(\theta',\lambda',R\right)\sin\alpha\right)$$
$$\frac{\partial}{\partial\psi}B(\psi,r)\mathrm{d}\sigma \qquad [97]$$

where

$$B(\psi,r) = \sum_{n=2}^{\infty}\frac{(2n+1)}{n(n+1)}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi) \qquad [98]$$

It is interesting to note that instead of an integral over the sphere, the inverse relationship between the disturbing potential on the sphere and the deflection of the vertical on the same sphere is also more straightforward in terms of a line integral:

$$T(\theta,\lambda,R) = T(\theta_0,\lambda_0,R) + \frac{\gamma_0}{4\pi}\int_{(\theta_0,\lambda_0)}^{(\theta,\lambda)}\left(\xi\left(\theta',\lambda',R\right)\mathrm{d}s_\theta-\eta\left(\theta',\lambda',R\right)\mathrm{d}s_\lambda\right) \qquad [99]$$

where

$$\mathrm{d}s_\theta = R\,\mathrm{d}\theta,\quad \mathrm{d}s_\lambda = R\sin\theta\ \mathrm{d}\lambda \qquad [100]$$

3.02.4.4 Concluding Remarks

The spherical harmonic series [65] represents the general solution to the exterior potential, regardless of the way the coefficients are determined. We know how to compute those coefficients exactly on the basis of a BVP, if the boundary is a sphere. More complicated boundaries would require corrections or, if these are omitted, would imply an approximation. If the coefficients are determined accurately (e.g., from satellite observations (**Section 3.02.6.1**), but not according to eqn [75]), then the spherical harmonic series model for the potential is not a spherical approximation. The spherical approximation enters when approximate relations such as eqn [5] are used and when the boundary is approximated as a sphere. However determined, the infinite series converges uniformly for all $r>R_c$, where R_c is the radius of the sphere that encloses all terrestrial masses (the *Brillouin sphere*). It may also converge in some sense below this sphere but would represent the true potential only in free space (above the Earth's surface, where Laplace's equation holds). In practice, though, convergence, per se, is not an issue since the series must be truncated at some finite degree. Any trend toward divergence is then part of the overall model error.

Model errors exist in all Green's function integrals as they depend on spherical approximations in the boundary condition. In addition, the surface of integration in these formulas is formally assumed to be the geoid (where normal derivatives of the potential coincide with gravity magnitude), but it is approximated as a sphere. The spherical approximation results, in the first place, from a neglect of the ellipsoid flattening, which is about 0.3% for the Earth. When working with the disturbing potential, this level of error was easily tolerated in the past (e.g., if the geoid undulation is 30 m, the spherical approximation accounts for about 10 cm error), but today, it requires attention as geoid undulation accuracy of 1 cm is pursued.

Green's functions all have singularities when the evaluation point is on the sphere of radius, R. For points above this sphere, it is easily verified that all the harmonic series for Green's functions converge uniformly, since $|P_n(\cos\psi)|\leq 1$. When $r=R$, the corresponding singularities of the integrals are either weak (Stokes' integral) or strong (e.g., Poisson's integral), requiring in the latter case special definition of the integral as a Cauchy principal value.

It is noted that the BVP solutions also require that no masses reside above the geoid (the boundary approximated as a sphere). To satisfy this condition, the topographic masses must be redistributed appropriately by mathematical reduction, and the gravity anomalies or disturbances measured on the Earth's surface must be reduced to the geoid. The mass redistribution must then be undone (mathematically), resulting in an indirect effect, in order to obtain the correct potential on or above the geoid. Details of these various procedures are found in Heiskanen and Moritz (1967, **Chapter 1.04**). The most common methodology today is the Helmert condensation method (Hofmann-Wellenhof and Moritz, 2005), whereby the topographic masses with assumed constant density are 'condensed' as a surface layer on the geoid, thus satisfying the conditions of the BVP. While this reduction has no particular geophysical interpretation, its indirect effect is generally smaller than for other reductions. The atmosphere, having significant mass, also affects gravity anomalies as a function of elevation. This effect may also be removed prior to using them as boundary values in the integral formulas (Moritz, 2000).

Finally, solutions to the BVP for the potential in *spheroidal* coordinates and ellipsoid boundary (see **Section 3.02.5.2**) replace the spherical approximation with a significantly improved ellipsoidal approximation, but there is no corresponding convolution theorem, as in eqn [59]. That is, integral solutions with analytic forms of a Green's function do not exist in this case because the inverse distance now depends on two surface coordinates. However, approximations have been formulated for all three types of BVP (see Yu et al., 2002 and references therein). In practice, solutions such as Stokes' integral (eqn [72]) still find general usage, especially if the

long-wavelength field is accommodated by a global model (remove–restore technique, Section 3.02.6.2) and where corrections are applied to reduce the spherical approximation error (e.g., Fei and Sideris, 2000).

3.02.5 Low-Degree Harmonics: Interpretation and Reference

The low-degree spherical harmonics of the Earth's gravitational potential lend themselves to interpretation with respect to the most elemental distribution of the Earth's density, which also leads to fundamental geometric characterizations, particularly for the second-degree harmonics. Let

$$C_{nm}^{(a)} = \frac{a}{GM}\left(\frac{R}{a}\right)^{n+1} v_{nm} \qquad [101]$$

be unit-less coefficients that refer to a sphere of radius, a. (Recall that coefficients, v_{nm} (eqn [50]), refer to a sphere of radius, R.) Relative to the central harmonic coefficient, $C_{00}^{(a)} = 1$, the next significant harmonic, $C_{20}^{(a)}$, for the Earth is more than three orders of magnitude smaller; and the remaining harmonic coefficients are at least two to three orders of magnitude smaller than that. The attenuation after degree two is much more gradual (Table 1), indicating that the bulk of the potential can be described by an ellipsoidal field. The normal gravitational field is such a field, but it also adheres to a geodetic definition that requires the underlying ellipsoid to be an equipotential surface in the corresponding normal gravity field. This section examines the low-degree harmonics from these two perspectives of interpretation and reference.

3.02.5.1 Low-Degree Harmonics as Density Moments

Returning to the general expression for the gravitational potential in terms of the Newtonian density integral (eqn [14]), and substituting the spherical harmonic series for the reciprocal distance (eqn [63] with [57]),

$$V(\theta,\lambda,r) = G\iiint_v \rho \frac{1}{r'}\sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^{n+1}\left(\frac{1}{2n+1}\sum_{m=-n}^{n}\overline{Y}_{nm}(\theta,\lambda)\overline{Y}_{nm}\left(\theta',\lambda'\right)\right)\mathrm{d}v$$
$$= \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+1}\left(\frac{G}{R^{n+1}(2n+1)}\iiint_{\text{volume}}\rho\left(r'\right)^{n}\overline{Y}_{nm}\left(\theta',\lambda'\right)\mathrm{d}v\right)$$
$$\overline{Y}_{nm}(\theta,\lambda) \qquad [102]$$

yields a multipole expansion (so called from electrostatics) of the potential. The spherical harmonic coefficients, also commonly called Stokes' constants or coefficients in this case, are multipoles of the density distribution (cf., eqn [50]):

$$v_{nm} = \frac{G}{R^{n+1}(2n+1)}\iiint_v \rho\left(r'\right)^{n}\overline{Y}_{nm}\left(\theta',\lambda'\right)\mathrm{d}v \qquad [103]$$

One may also consider the nth-order moments of density (from the statistics of distributions) defined by

$$\mu_{\alpha\beta\gamma}^{(n)} = \iiint_v \left(x'\right)^{\alpha}\left(y'\right)^{\beta}\left(z'\right)^{\gamma}\rho\,\mathrm{d}v, \quad n = \alpha+\beta+\gamma \qquad [104]$$

(where, clearly risking confusion, we defer to the common nomenclature of order for the moments and degree for the spherical harmonics. Also for notational convenience, here, we use (x,y,z) for the Cartesian coordinates instead of (x_1, x_2, x_3) used in previous sections.) The multipoles of degree n and the moments of order n are related, although not all $(n+1)(n+2)/2$ moments of order n can be determined from the $2n+1$ multipoles of degree n, when $n \geq 2$. This indeterminacy is directly connected to the inability to determine the density distribution uniquely from external measurements of the potential (Chao, 2005), which is the classic geophysical inverse problem.

The zero-degree Stokes' coefficient is coordinate-invariant and is proportional to the total mass of the Earth:

$$v_{00} = \frac{G}{R}\iiint_v \rho\,\mathrm{d}v = \frac{GM}{R} \qquad [105]$$

and, it also represents a mass monopole, and it is proportional to the zeroth moment of the density, M.

The first-degree harmonic coefficients (representing dipoles) are proportional to the coordinates of the center of mass, (x_{cm},y_{cm},z_{cm}), which are proportional to the first-order moments of the density, as verified by recalling the definition of the first-degree spherical harmonics:

$$v_{1m} = \frac{G}{\sqrt{3}R^2}\iiint_v \rho\begin{Bmatrix} r'\sin\theta'\sin\lambda', & m=-1 \\ r'\cos\theta', & m=0 \\ r'\sin\theta'\cos\lambda', & m=1 \end{Bmatrix}\mathrm{d}v$$
$$= \frac{GM}{\sqrt{3}R^2}\begin{Bmatrix} y_{cm}, & m=-1 \\ z_{cm}, & m=0 \\ x_{cm}, & m=1 \end{Bmatrix} \qquad [106]$$

Nowadays, by tracking satellites, we have access to the center of mass of the Earth (including its atmosphere) since it defines the center of their orbits. Ignoring the small motion of the center of mass (annual amplitude of several millimeters) due to the temporal variations in the mass distribution, we may choose the coordinate origin for the geopotential model to coincide with the center of mass, thus annihilating the first-degree coefficients.

Table 1 Spherical harmonic coefficients of the total gravitational potential[a]

Degree, n	*Order*, m	$C_{nm}^{(a)}$	$C_{n,-m}^{(a)}$
2	0	−4.84170E−04	0.0
2	1	−2.39832E−10	1.42489E−09
2	2	2.43932E−06	−1.40028E−06
3	0	9.57189E−07	0.0
3	1	2.03048E−06	2.48172E−07
3	2	9.04802E−07	−6.19006E−07
3	3	7.21294E−07	1.41437E−06
4	0	5.39992E−07	0.0
4	1	−5.36167E−07	−4.73573E−07
4	2	3.50512E−07	6.62445E−07
4	3	9.90868E−07	−2.00976E−07
4	4	−1.88472E−07	3.08827E−07

[a]GRACE model GGM02S (Tapley et al., 2005).

The second-order density moments likewise are related to the second-degree harmonic coefficients (quadrupoles). They also define the inertia tensor of the body. The inertia tensor is the proportionality factor in the equation that relates the angular momentum vector, H, and the angular velocity, $\boldsymbol{\omega}$, of a body, like the Earth

$$\mathrm{H} = I\boldsymbol{\omega} \qquad [107]$$

and is denoted by

$$I = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \qquad [108]$$

It comprises the moments of inertia on the diagonal

$$\begin{aligned} I_{xx} &= \iiint_v \rho\left(y'^2 + z'^2\right) \mathrm{d}v, \quad I_{yy} \\ &= \iiint_v \rho\left(z'^2 + x'^2\right) \mathrm{d}v, \quad I_{zz} = \iiint_v \rho\left(x'^2 + y'^2\right) \mathrm{d}v \end{aligned} \qquad [109]$$

and the products of inertia off the diagonal

$$\begin{aligned} I_{xy} = I_{yx} &= -\iiint_v \rho\ x'y'\, \mathrm{d}v, \quad I_{xz} = I_{zx} \\ &= -\iiint_v \rho\ x'z'\, \mathrm{d}v, \quad I_{yz} = I_{zy} = -\iiint_v \rho\ y'z'\, \mathrm{d}v \end{aligned} \qquad [110]$$

Note that

$$I_{xx} = \mu^{(2)}_{020} + \mu^{(2)}_{002}, \quad I_{xy} = -\mu^{(2)}_{110} \text{ etc.}, \qquad [111]$$

and there are as many (six) independent tensor components as second-order density moments. Using the explicit expressions for the second-degree spherical harmonics, we have from eqn [103] with $n=2$

$$\begin{aligned} &v_{2,-2} = -\frac{\sqrt{15}\,G}{5R^3} I_{xy}, \quad v_{2,-1} = -\frac{\sqrt{15}\,G}{5R^3} I_{yz}, \quad v_{2,1} = -\frac{\sqrt{15}\,G}{5R^3} I_{xz} \\ &v_{2,0} = \frac{\sqrt{5}\,G}{10R^3}\left(I_{xx} + I_{yy} - 2I_{zz}\right), \quad v_{2,2} = \frac{\sqrt{15}\,G}{10R^3}\left(I_{yy} - I_{xx}\right) \end{aligned} \qquad [112]$$

which are also known as *MacCullagh's* formulas. Not all density moments (or moments of inertia) can be determined from these five Stokes' coefficients.

If the coordinate axes are chosen so as to diagonalize the inertia tensor (products of inertia are then equal to zero), then they are known as principal axes of inertia, or also 'figure' axes. For the Earth, the z-figure axis is very close to the spin axis (within several meters at the pole); both axes move with respect to each other and the Earth's surface (known as polar motion), with combinations of various periods (daily, monthly, annually, etc.) and secularly in a wandering fashion. Because of these motions, the figure axis is not useful as a coordinate axis that defines a frame fixed to the crust of the Earth. However, because of the proximity of the figure axis to the defined reference z-axis, the second-degree, first-order harmonic coefficients of the geopotential are relatively small (Table 1).

The arbitrary choice of the x-axis of our Earth-fixed reference coordinate system certainly did not attempt to eliminate the product of inertia, I_{xy} (the x-axis is defined by the intersection of the Greenwich meridian with the equator; and the y-axis completes a right-handed mutually orthogonal triad). However, it is possible to determine where the equatorial figure axis is located by combining values of the second-degree, second-order harmonic coefficients. Let u, v, and w be the axes that define a coordinate system in which the inertia tensor is diagonal; and assume that $I_{ww} = I_{zz}$. Let λ_0 be the longitude of the u-figure axis in our conventional (x,y,z) system. A rotation by the angle, $-\lambda_0$, about the w-(also z-)figure axis brings this ideal coordinate system back to the conventional one in which we calculate the harmonic coefficients. Tensors transform under rotation, defined by matrix, $\mathcal{R}$, according to

$$I_{xyz} = \mathcal{R}\, I_{uvw} \mathcal{R}^{\mathrm{T}} \qquad [113]$$

where the subscript on I denotes the coordinate system to which it refers. With the rotation about the w-axis given by the matrix

$$\mathcal{R} = \begin{pmatrix} \cos\lambda_0 & -\sin\lambda_0 & 0 \\ \sin\lambda_0 & \cos\lambda_0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad [114]$$

and with eqn [112], it is straightforward to show that

$$\begin{aligned} v_{2,-2} &= -\frac{\sqrt{15}\,G}{10R^3}\left(I_{uu} - I_{vv}\right) \sin 2\lambda_0 \\ v_{2,2} &= -\frac{\sqrt{15}\,G}{10R^3}\left(I_{uu} - I_{vv}\right) \cos 2\lambda_0 \end{aligned} \qquad [115]$$

Hence, we have

$$\lambda_0 = \frac{1}{2} \tan^{-1} \frac{v_{2,-2}}{v_{2,2}} \qquad [116]$$

where the quadrant is determined by the signs of the harmonic coefficients. From Table 1, we find that $\lambda_0 = -14.929°$; that is, the u-figure axis is in the mid-Atlantic between South America and Africa.

The second-degree, second-order harmonic coefficient, $v_{2,2}$, indicates the asymmetry of the Earth's mass distribution with respect to the equator. Since $v_{2,2} > 0$ (for the Earth), eqn [115] shows that $I_{vv} > I_{uu}$ and thus the equator 'bulges' more in the direction of the u-figure axis; conversely, the equator is flattened in the direction of the v-figure axis. This flattening is relatively small: 1.1×10^{-5}.

Finally, consider the most important second-degree harmonic coefficient, the second zonal harmonic, $v_{2,0}$. Irrespective of the x-axis definition, it is proportional to the difference between the moment of inertia, I_{zz}, and the average of the equatorial moments, $(I_{xx} + I_{yy})/2$. Again, since $v_{2,0} < 0$, the Earth bulges more around the equator and is flattened at the poles. The second zonal harmonic coefficient is roughly 1000 times larger than the other second-degree coefficients and thus indicates a substantial polar flattening (owing to the Earth's early more fluid state). This flattening is approximately 0.003.

3.02.5.2 Normal Ellipsoidal Field

Because of Earth's dominant polar flattening and the near symmetry of the equator, any meridional section of the Earth

is closer to an ellipse than a circle. For this reason, spheroidal coordinates (eqn [2]) have often been advocated in place of the usual spherical coordinates. In fact, for geodetic positioning and geographic mapping, because of this flattening, the conventional (geodetic) latitude and longitude define the direction of the perpendicular to an ellipsoid, rather than the radial direction from the origin. These geodetic coordinates, however, are not the spheroidal (ellipsoidal) coordinates defined in **Section 3.02.1** and would be rather useless in potential modeling because they do not separate Laplace's differential equation.

Solutions to Laplace's equation in terms of the spheroidal coordinates, (δ,λ,u), are developed easily, analogous to the spherical harmonic series (**Section 3.02.4**) (Hofmann-Wellenhof and Moritz, 2005). Until recently, they have not been adopted in most geodetic applications perhaps in part because of the nonintuitive nature of the coordinates. On the other hand, it is definitely advantageous to model the normal (or reference) gravity field in terms of these coordinates since it is based on an ellipsoid and the corresponding simplicity of the boundary condition results in closed formulas for the potential. Separating Laplace's equation analogous to its form in spherical coordinates, the solution is obtained by successively solving three ordinary differential equations. Applied to the exterior gravitational potential, V, the solution is given by (cf. eqn [50])

$$V(\delta,\lambda,u)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{Q_{n|m|}(iu/E)}{Q_{n|m|}(ib/E)}v_{nm}^{e}\overline{Y}_{nm}(\delta,\lambda) \qquad [117]$$

where E is the linear eccentricity associated with the coordinate system and Q_{nm} is the associated Legendre function of the second kind. The coefficients of the series, v_{nm}^{e}, refer to an ellipsoid of semiminor axis, b; and with the series written in this way, they are all real numbers with the same units as V. An exact relationship between these and the spherical harmonic coefficients, v_{nm}, was given by Hotine (1969) and Jekeli (1988). The computation of the functions, Q_{nm}, was advanced recently by Sebera et al. (2012).

With this formulation of the potential, Dirichlet's BVP is solved for an ellipsoidal boundary using the orthogonality of the spherical harmonics:

$$v_{nm}^{e}=\frac{1}{4\pi}\iint_{\sigma}V(\delta,\lambda,b)\overline{Y}_{nm}(\delta,\lambda)\,\mathrm{d}\sigma \qquad [118]$$

where $\mathrm{d}\sigma=\sin\delta\,\mathrm{d}\delta\,\mathrm{d}\lambda$ and $\sigma=\{(\delta,\lambda)|0\le\delta\le\pi,\ 0\le\lambda\le2\pi\}$. Note that while the limits of integration and the differential element, $\mathrm{d}\sigma$, are the same as for the unit sphere, the boundary values are on the ellipsoid.

The simplicity of the boundary values of the *normal* gravitational potential allows its extension into exterior space to be expressed in closed analytic form. Analogous to the geoid in the actual gravity field, the normal ellipsoid is defined to be a level surface in the normal *gravity* field. In other words, the sum of the normal gravitational potential and the centrifugal potential due to Earth's rotation is a constant on the ellipsoid:

$$V^{e}(\delta,\lambda,b)+\phi(\delta,b)=U_0 \qquad [119]$$

Hence, the normal gravitational potential on the ellipsoid, $V^{e}(\delta,\lambda,b)$, depends only on latitude and is symmetrical with respect to the equator. Consequently, it consists of only even zonal harmonics, and because the centrifugal potential has only zero- and second-degree zonals, the corresponding ellipsoidal series is finite (up to degree two). The solution to this Dirichlet problem is given in spheroidal coordinates by

$$V^{e}(\delta,\lambda,u)=\frac{GM}{E}\tan^{-1}\frac{E}{u}+\frac{1}{2}\omega_{e}^{2}a^{2}\frac{q}{q_0}\left(\cos^{2}\delta-\frac{1}{3}\right) \qquad [120]$$

where a is the semimajor axis of the ellipsoid, ω_e is Earth's rate of rotation, and

$$q=\frac{1}{2}\left(\left(1+3\frac{u^2}{E^2}\right)\tan^{-1}\frac{E}{u}-3\frac{u}{E}\right),\quad q_0=q|_{u=b} \qquad [121]$$

Heiskanen and Moritz (1967) and Hofmann-Wellenhof and Moritz (2005) provided details of the straightforward derivation of these and the following expressions.

The equivalent form of V^{e} in spherical harmonics is given by

$$V^{e}(\theta,\lambda,r)=\frac{GM}{r}\left(1-\sum_{n=1}^{\infty}J_{2n}\left(\frac{a}{r}\right)^{2n}P_{2n}(\cos\theta)\right) \qquad [122]$$

where

$$J_{2n}=(-1)^{n+1}\frac{3e^{2n}}{(2n+1)(2n+3)}\left(1-n+\frac{5n}{e^2}J_2\right),\quad n\ge1 \qquad [123]$$

and $e=E/a$ is the first eccentricity of the ellipsoid. The second zonal coefficient is given by

$$J_2=\frac{e^2}{3}\left(1-\frac{2}{15q_0}\frac{\omega_e^2a^2E}{GM}\right)=\frac{I_{zz}^{e}-I_{xx}^{e}}{Ma^2} \qquad [124]$$

where the second equality comes directly from the last of eqn [112] and the ellipsoid's rotational symmetry ($I_{xx}^{e}=I_{yy}^{e}$). This equation also provides a direct relationship between the geometry (the eccentricity or flattening) and the mass distribution (difference of second-order moments) of the ellipsoid. Therefore, J_2 is also known as the dynamic form factor – the flattening of the ellipsoid can be described either geometrically or dynamically in terms of a difference in density moments.

The normal gravitational potential depends solely on four adopted parameters: Earth's rotation rate, ω_e; the size and shape of the normal ellipsoid, for example, a and J_2; and a potential scale, for example, GM. The mean Earth ellipsoid is the normal ellipsoid with parameters closest to actual corresponding parameters for the Earth. GM and J_2 are determined by observing satellite orbits, a can be calculated by fitting the ellipsoid to mean sea level using satellite altimetry, and Earth's rotation rate comes from astronomical observations. **Table 2** gives present values and adopted constants for the Geodetic Reference Systems of 1967 and 1980 (GRS67 and GRS80) and the World Geodetic System 1984 (WGS84).

In modeling the disturbing potential in terms of spherical harmonics, one naturally uses the form of the normal gravitational potential given by eqn [122]. Here, we have assumed that all harmonic coefficients refer to a sphere of radius, a. Corresponding coefficients for the series of $T=V-V^{e}$ are, therefore,

Table 2 Defining parameters for normal ellipsoids of geodetic reference systems

Reference system	a *(m)*	J_2	GM *($m^3\ s^{-2}$)*	ω_e *(rad s^{-2})*
GRS67	6378160	1.0827E−3	3.98603E14	7.2921151467E−5
GRS80	6378137	1.08263E−3	3.986005E14	7.292115E−5
WGS84	6378137	1.08262982131E−3	3.986004418E14	7.2921151467E−5
Current IERS values[a]	6378136.6 ± 0.1 (mean–tide system)	(1.0826359 ± 0.0000001) E−03 (zero-tide system)	(3.986004418 ± 0.000000008)E14 (includes atmosphere)	7.292115E−5 (mean value)

[a]Adapted by the International Earth Rotation and Reference System Service (IERS, Petit and Luzum, 2010).

$$\delta C_{nm}^{(a)} = \begin{cases} 0, & n = 0 \\ C_{n0}^{(a)} - \dfrac{-J_n}{\sqrt{2n+1}}, & n = 2, 4, 6, \ldots \\ C_{nm}^{(a)}, & \text{otherwise} \end{cases} \qquad [125]$$

For coefficients referring to the sphere of radius, R, as in eqn [65], we have $\delta C_{nm} = (a/R)^n \delta C_{nm}^{(a)}$. The harmonic coefficients, J_{2n}, attenuate rapidly due to the factor, e^{2n}; and only harmonics up to degree ten are significant.

Normal gravity, being the gradient of the normal gravity potential, is used typically only in applications tied to an Earth-fixed coordinate system. Exact formulas for all its components on and above the ellipsoid may be found in the WGS84 document (NIMA, 2000). For points on the ellipsoid, the only nonzero component is along the ellipsoid normal and is given by the Somigliana formula:

$$\gamma(\phi) = \gamma_a \frac{1 + \left(\frac{b\gamma_b}{a\gamma_a} - 1\right) \sin^2\phi}{\sqrt{1 - e^2 \sin^2\phi}} \qquad [126]$$

where ϕ is geodetic latitude, a, b are the semimajor and semiminor axes of the ellipsoid, and γ_a, γ_b are the values of normal gravity, respectively, at the equator and pole of the ellipsoid. The values of the latter (and b) are constants derived exactly from the four fundamental constants of the normal ellipsoid (Moritz, 2000).

3.02.6 Methods of Determination

In this section, we briefly explore the basic technologies that yield measurements of the gravitational field. Even though we have reduced the problem of determining the exterior potential from a volume integral to a surface integral (e.g., either eqn [72] or [75]), it is clear that in theory, we can never determine the entire field from a finite number of measurements. The integrals will always need to be approximated numerically, and/or the infinite series of spherical harmonics needs to be truncated. However, with enough effort and within the limits of computational capabilities, one can approach the ideal continuum of boundary values as closely as desired or make the number of coefficients in the series representation as large as possible. The expended computational and measurement efforts have to be balanced with the ability to account for inherent model errors (such as the spherical approximation) and the noise of the measuring device. To be useful for geodetic and geodynamic purposes, the instruments must possess a sensitivity of at least a few parts per million, and, in fact, many have a sensitivity of parts per billion. These sensitivities often come at the expense of prolonging the measurements (integration time) in order to average out random noise, thus reducing the achievable temporal resolution. This is particularly critical for moving-base instrumentation such as on an aircraft or satellites where temporal resolution translates into spatial resolution through the velocity of the vehicle.

3.02.6.1 Measurement Systems and Techniques

Determining the gravitational field through classical measurements relies on three fundamental laws. The first two are Newton's second law of motion and his law of gravitation. Newton's law of motion states that the rate of change of linear momentum of a particle is equal to the totality of forces, $\boldsymbol{F}$, acting on it. Given more familiarly as $m_i \mathrm{d}^2\boldsymbol{x}/\mathrm{d}t^2 = \boldsymbol{F}$, it involves the inertial mass, m_i; and, conceptually, the forces, $\boldsymbol{F}$, should be interpreted as action forces, like propulsion or friction. The gravitational field, which is part of the space we occupy, is due to the presence of masses like the Earth, sun, moon, and planets and induces a different kind of force, the gravitational force (**Section 3.02.2**). It is related to gravitational acceleration, $\boldsymbol{g}$, through the gravitational mass, m_g, according to the law of gravitation, abbreviated here as $m_g \boldsymbol{g} = \boldsymbol{F}_g$. Newton's law of motion must be modified to include $\boldsymbol{F}_g$ separately. Through the third fundamental law, Einstein's equivalence principle, which states that inertial and gravitational masses are indistinguishable, we finally get

$$\frac{\mathrm{d}^2\boldsymbol{x}}{\mathrm{d}t^2} = \boldsymbol{a} + \boldsymbol{g} \qquad [127]$$

where $\boldsymbol{a}$ is the specific force ($\boldsymbol{F}/m_i$), or also the inertial acceleration, due to action forces. This equation holds in a nonrotating, freely falling frame (i.e., an inertial frame), and variants of it can be derived in more complicated frames that rotate or have their own dynamic motion. However, one can always assume the existence of an inertial frame and proceed on that basis.

There exist a variety of devices that measure the motion of an inertial mass with respect to the frame of the device, and thus, technically, they sense $\boldsymbol{a}$; such devices are called accelerometers. Consider the special case that an accelerometer is resting on the Earth's surface with its sensitive axis aligned along the vertical. In an Earth-centered frame (inertial, if we ignore Earth's rotation), the free-fall motion of the accelerometer is impeded by the reaction force of the Earth's surface acting on the accelerometer. In this case, the left side of eqn [127] applied to the motion of the accelerometer is zero,

and the accelerometer, sensing the reaction force, indirectly measures (the negative of) gravitational acceleration. This accelerometer is given the special name, gravimeter.

Gravimeters, especially static instruments, are designed to measure acceleration at very low temporal frequencies (i.e., averaged over longer periods of time), whereas accelerometers, typically used in navigation or other motion-sensing applications, require a much shorter response in order to sense the rapidly changing accelerations. As such, gravimeters generally are more accurate. Earth-fixed gravimeters actually measure gravity (**Section 3.02.1.3**), the combination of gravitation and centrifugal acceleration due to Earth's spin (the frame is not inertial in this case). The simplest, though not the first invented, gravimeter utilizes a vertically, freely falling mass, measuring the time it takes to fall a given distance. Applying eqn [127] to the falling mass in the frame of the device, one can solve for g (assuming it is constant): $x(t) = 0.5gt^2$. Equations can also be developed that account for the linear gradient of gravity. This free-fall gravimeter is a special case of a more general gravimeter that constrains the fall using an attached spring or the arm of a pendulum, where other (action) forces (the tension in the arm or the spring) thus enter into the equation.

The first gravimeter, in fact, was the pendulum, systematically used for gravimetry as early as the 1730s and 1740s by P. Bouguer on a geodetic expedition to measure the size and shape of the Earth (meridian arc measurement in Peru). The pendulum served well into the twentieth century (until the early 1970s) both as an absolute device, measuring the total gravity at a point, and as a relative device indicating the difference in gravity between two points (Torge, 1989). Today, absolute gravimeters exclusively rely on a freely falling mass, where exquisitely accurate measurements of distance and time are achieved with laser interferometers and atomic clocks (Niebauer et al., 1995; Zumberge et al., 1982). Accurate *relative* gravimeters are much less expensive, requiring a measurement of distance change only, and because many errors that cancel between measurements need not be addressed. They rely almost exclusively on a spring-suspended test mass (Nettleton, 1976; Torge, 1989). Developed early in the twentieth century in response to oil exploration requirements, the basic principles of the relative gravimeter have changed little since then. Modern instruments include electronic recording capability and specialized stabilization and damping for deployment on moving vehicles such as ships and aircraft. The accuracy of absolute gravimeters is typically of the order of parts per billion, and relative devices in field deployments may be as good but more typically are at least one order of magnitude less precise. Laboratory relative (in time) gravimeters, based on cryogenic instruments that monitor the virtual motion of a test mass that is electromagnetically suspended using superconducting persistent currents (Goodkind, 1999), are as accurate as portable absolute devices (or more), owing to the stability of the currents and the controlled laboratory environment.

On a moving vehicle, particularly an aircraft, the relative gravitational acceleration can be determined only from a combination of gravimeter and kinematic positioning system. The latter is needed to derive the (vertical) kinematic acceleration (the left side of eqn [127]). Today, the GPS best serves that function, yielding centimeter-level precision in relative three-dimensional position. Such combined GPS/gravimeter systems have been used successfully to determine the vertical component of gravitation over large otherwise inaccessible areas such as the Arctic Ocean (Kenyon and Forsberg, 2000), as well as in regions where conventional surface gravimetry would be much less economical for geoid determination and oil exploration purposes (Gumert, 1998; Hammer, 1983; Olesen and Forsberg, 2007). The airborne gravimeter is specially designed to damp out high-frequency noise and usually is stabilized on a level platform.

Three-dimensional moving-base gravimetry has also been demonstrated using the triad of accelerometers of a high-accuracy inertial navigation system (the type that are fixed to the aircraft without special stabilizing platforms). The orientation of all accelerometers on the vehicle must be known with respect to inertial space, which is accomplished with precision gyroscopes. Again, the total acceleration vector of the vehicle, $\mathrm{d}^2\boldsymbol{x}/\mathrm{d}t^2$, can be ascertained by time differentiation of the kinematic positions (from GPS). One of the most critical errors is due to the cross coupling of the horizontal orientation error, $\delta\psi$, with the large vertical acceleration (the lift of the aircraft, essentially equal to $-g$). This is a first-order effect ($g \sin \delta\psi$) in the estimation of the horizontal gravitation components but only a second-order effect ($g(1-\cos\delta\psi)$) on the vertical component. Details of moving-base vector gravimetry may be found in Jekeli (2000a, **Chapter 1.11**), Kwon and Jekeli (2001), and also Studinger et al. (2008). Further developments and examples of geophysical applications may be found in Lane (2004, 2010).

The ultimate global gravimeter is the satellite in free fall (i.e., in orbit due to sufficient forward velocity) – the satellite is the inertial mass and the 'device' is a set of reference points with known coordinates, for example, on the Earth's surface, or another satellite whose orbit is known. The measuring technology is an accurate ranging system (radar or laser) that tracks the satellite as it orbits (falls to) the Earth. Ever since Sputnik, the first artificial satellite launched into Earth orbit in 1957, Earth's gravitational field could be determined by tracking satellites from precisely known ground stations. Equation [127], with gravitational acceleration expressed as a truncated series of spherical harmonics (gradient of eqn [50]), becomes

$$\frac{\mathrm{d}^2\boldsymbol{x}}{\mathrm{d}t^2} = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} v_{nm} \nabla \left(\left(\frac{R}{r} \right)^{n+1} Y_{nm}(\theta < \lambda + \omega_e t) \right) + \delta \boldsymbol{R} \quad [128]$$

where ω_e is Earth's rate of rotation and $\delta\boldsymbol{R}$ represents residual accelerations due to action forces (solar radiation pressure, atmospheric drag, Earth's albedo, etc.), gravitational tidal accelerations due to other bodies (moon, sun, and planets), and all other subsequent indirect effects. The position vector on the left side of eqn [128] is more explicitly $\boldsymbol{x}(t) = \boldsymbol{x}(\theta(t), \lambda(t) + \omega_e t, r(t))$; and the spatial coordinates on the right side are also functions of time. This makes the equation more conceptual than practical since it is numerically more convenient to transform the satellite position and velocity into Keplerian orbital elements (semimajor axis of the orbital ellipse, its eccentricity, its inclination to the equator, the angle of perigee, the right ascension of the node of the orbit, and the mean motion), all of which also change in time, but most much

more slowly. This transformation was derived by Kaula (1966); see also Seeber (1993).

In the most general case ($n_{\max} > 2$ and $\delta \boldsymbol{R} \neq 0$), there is no analytic solution to eqn [128] or to its transformations to other types of coordinates. The positions of the satellite are observed by ranging techniques, and the unknowns to be solved are the coefficients, v_{nm}. Numerical integration algorithms have been specifically adapted to this problem, and extremely sophisticated models for $\delta \boldsymbol{R}$ are employed with additional unknown parameters to be solved in order to estimate as accurately as possible the gravitational coefficients (e.g., Cappelari et al., 1976; Pavlis et al., 1999). The entire procedure falls under the broad category of dynamic orbit determination, and the corresponding gravitational field modeling may be classified as the *time-wise* approach. The partial derivatives of eqn [128] with respect to unknown parameters, $\boldsymbol{p} = \{\ldots, v_{nm}, \ldots\}$, termed the *variational equations*, are integrated numerically in time, yielding estimates for $H = \partial \boldsymbol{x}/\partial \boldsymbol{p} = \{\ldots, \partial \boldsymbol{x}/\partial v_{nm}, \ldots\}$. These are then used in a least-squares adjustment of the linearized model relating observed positions (e.g., via ranges) to parameters:

$$\delta \boldsymbol{x} = H \delta \boldsymbol{p} + \varepsilon \qquad [129]$$

where $\delta \boldsymbol{x}$ and $\delta \boldsymbol{p}$ are differences with respect to previous estimates and $\boldsymbol{\varepsilon}$ represents errors (Tapley, 1973).

The alternative to this rather complicated estimation of the gravitational field by tracking satellites is the *in situ* measurement of the field. However, a gravimeter (or accelerometer) on a satellite does not sense the presence of a gravitational field. This is evident from the fact that the satellite is in free fall (apart from small accelerations due to action forces such as atmospheric drag) and the inertial test mass and the gravimeter, itself, are equally affected by gravitation (i.e., they are all in free fall). On the other hand, two accelerometers fixed on a satellite yield, through the difference in their outputs, a gradient in acceleration that includes the gradient of gravitation. On a nonrotating satellite, the acceleration at an arbitrary point, $\boldsymbol{b}$, of the satellite is given by

$$\boldsymbol{a}_b = \frac{\mathrm{d}^2 \boldsymbol{x}}{\mathrm{d}t^2} - \boldsymbol{g}(\boldsymbol{b}) \qquad [130]$$

in a coordinate system with origin at the center of mass of the satellite. Taking the difference (differential) of two accelerations in ratio to their separation, $\delta \boldsymbol{b}$, we obtain

$$\frac{\delta \boldsymbol{a}_b}{\delta \boldsymbol{b}} = -\frac{\delta \boldsymbol{g}}{\delta \boldsymbol{b}} \qquad [131]$$

where the ratios represent tensors of derivatives in the local satellite coordinate frame. That is, the kinematic acceleration of the satellite is the same for both points on the satellite and cancels. For a rotating satellite, this equation generalizes to

$$\frac{\partial \boldsymbol{a}_b}{\partial \boldsymbol{b}} = -\frac{\partial \boldsymbol{g}_b}{\partial \boldsymbol{b}} + \Omega^2 + \frac{\mathrm{d}}{\mathrm{d}t}\Omega \qquad [132]$$

where Ω is a skew-symmetrical matrix whose off-diagonal elements are the components of the vector that defines the rotation rate of the satellite with respect to the inertial frame.

Thus, a *gravitational gradiometer* on a satellite (or any moving vehicle) senses a combination of gravitational gradient and angular acceleration (including a centrifugal type). In principle, if the entire tensor of gradients is measured, then, because of the symmetry of the gravitational gradient tensor and of Ω^2 and the antisymmetry of $\mathrm{d}\Omega/\mathrm{d}t$, the sum $\delta \boldsymbol{a}_b/\delta \boldsymbol{b} + (\delta \boldsymbol{a}_b/\delta \boldsymbol{b})^{\mathrm{T}}$ eliminates the latter, while the difference $\delta \boldsymbol{a}_b/\delta \boldsymbol{b} - (\delta \boldsymbol{a}_b/\delta \boldsymbol{b})^{\mathrm{T}}$ can be used to infer Ω, subject to initial conditions.

When the two ends of the gradiometer are fixed to one frame, the common linear acceleration, $\mathrm{d}^2\boldsymbol{x}/\mathrm{d}t^2$, cancels, as shown earlier; but if the two ends are independent, disconnected platforms moving in similar orbits, the gravitational difference depends also on their relative motion in inertial space. This is the concept for satellite-to-satellite tracking, by which one satellite precisely tracks the other and the change in the range rate between them is the consequence of a combination of gravitational difference, a difference in action forces, and a centrifugal acceleration due to the rotation of the baseline of the satellite pair. It can be shown that the line-of-sight acceleration (the measurement) is given by

$$\frac{\mathrm{d}^2 \rho}{\mathrm{d}t^2} = \boldsymbol{e}_\rho^{\mathrm{T}}(\boldsymbol{g}(\mathrm{x}_2) - \boldsymbol{g}(\mathrm{x}_1)) + \boldsymbol{e}_\rho^{\mathrm{T}}(\boldsymbol{a}_2 - \boldsymbol{a}_1) + \frac{1}{\rho}\left(\left|\frac{\mathrm{d}}{\mathrm{d}t}\Delta \mathrm{x}\right|^2 - \left(\frac{\mathrm{d}\rho}{\mathrm{d}t}\right)^2\right) \qquad [133]$$

where $\boldsymbol{e}_\rho$ is the unit vector along the instantaneous baseline connecting the two satellites, ρ is the baseline length (the range), and $\Delta \boldsymbol{x} = \boldsymbol{x}_2 - \boldsymbol{x}_1$ is the difference in position vectors of the two satellites. Clearly, only the gravitational difference projected along the baseline can be determined (similar to a single-axis gradiometer) and then only if both satellites carry accelerometers that sense the nongravitational accelerations. Also, the orbits of both satellites must be known in order to account for the last (centrifugal) term.

An Earth-orbiting satellite clearly is the ideal platform on which to measure the gravitational field when seeking global coverage in relatively short time. One simply designs the orbit to be polar and circular; and as the satellite orbits, the Earth spins underneath, offering a different section of its surface on each satellite revolution. There are also limitations. The satellite must have low altitude, h, above the Earth's sphere of radius, R, in order to achieve high sensitivity, since the nth-degree harmonics of the field attenuate as $(R/(R+h))^{n+1}$. On the other hand, the lower the altitude, the shorter is the life of the satellite due to atmospheric drag, which can only be countered with onboard propulsion systems. Secondly, because of the inherent speed of lower-orbit satellites (about 7 km s^{-1}), the resolution of its measurements is limited by the integration (averaging) time of the sensor (typically 1–10 s). Higher resolution comes only with shorter integration time, which may reduce the accuracy if this depends on averaging out random noise. Figure 9 shows the corresponding achievable resolution on the Earth's surface for different satellite instrumentation parameters, length of time in polar orbit and along-orbit integration time, or smoothing (Jekeli, 2004). In each case, the indicated level of resolution is warranted only if the noise of the sensor (after smoothing) does not overpower the signal at this resolution.

Being a higher-order derivative of the gravitational potential, gravitational gradients are particularly revealing of sharp contrasts in Earth's mass density structure. And although the gravitational gradients attenuate away from a source body with one additional power of distance (compared to gravitational

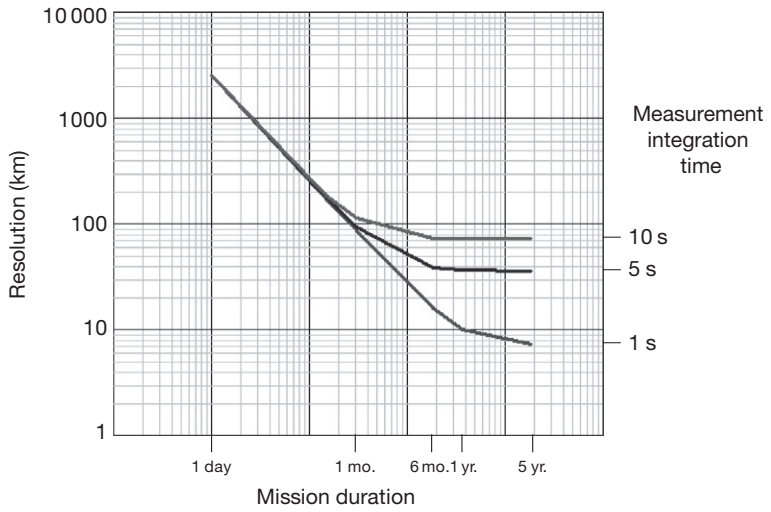


Figure 9 Spatial resolution of satellite measurements versus mission duration and integration time. The satellite altitude is 450 km. Reproduced from Jekeli C (2004) High-resolution gravity mapping: the next generation of sensors. In: Sparks RSJ (ed.) The State of the Planet: Frontiers and Challenges in Geophysics, Geophysical Monograph 150, IUGG, vol. 19, pp. 135–146, with permission from American Geophysical Union.

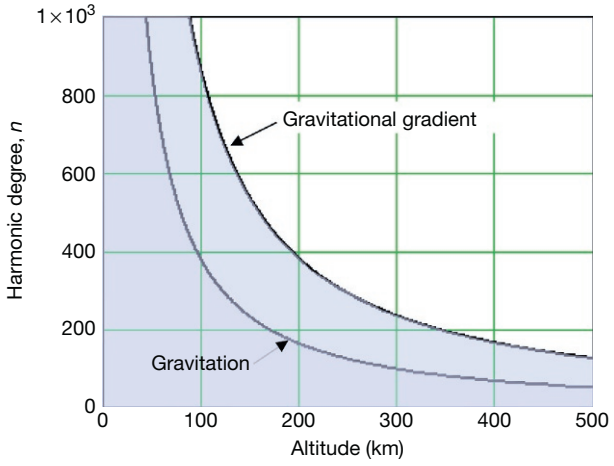


Figure 10 The product, $n^q(R/(R+h))^{n+1+q}$, of derivative and upward continuation frequency responses as a function of altitude, h, and harmonic degree, n, for gravitation ($q=1$) and its gradient ($q=2$). The shaded area represents the domain in each case where this product is greater than unity and thus does not attenuate the field.

acceleration), this is more than compensated by the increased sensitivity to (shorter) wavelengths. Indeed, the geopotential field spectrum scales by $\sim n^q$ due to q radial derivatives of the field, while it scales by $(R/(R+h))^{n+1+q}$ due to upward continuation. When the combination of these effects is greater than unity, there is no physical attenuation in the corresponding spectral component. As seen in Figure 10, for all altitudes, the recoverable spectrum generally is always greater with gravity gradiometry ($q=2$) than with gravimetry ($q=1$). Thus, for optimal resolution, *in situ* gravity gradiometry is preferred on satellite missions.

Two satellite-to-satellite tracking systems were launched to determine the global gravitational field. One is CHAMP (Challenging Minisatellite Payload) in 2000 (Reigber et al., 2002) and the other is GRACE (Gravity Recovery and Climate Experiment) in 2002 (Tapley et al., 2004a). CHAMP is a single low-orbiting satellite (400–450 km altitude) being tracked by the high-altitude GPS satellites, and it also carries a magnetometer to map the Earth's magnetic field. GRACE is more specifically dedicated to determining with extremely high accuracy the long to medium wavelengths of the gravitational field and their temporal variations. With two satellites in virtually identical low Earth orbits, one following the other, the primary GRACE data consist of intersatellite ranges observed by K/Ka-band radar with 10 μm precision in the 10 Hz bandwidth. The objective is to sense changes in the gravitational field due to mass transfer on the Earth within and among the atmosphere, the hydrosphere/cryosphere, and the oceans (Tapley et al., 2004b).

A gravity gradiometer was launched for the first time in March 2009 with the satellite GOCE (Gravity Field and Steady-State Ocean Circulation Explorer; Rummel et al., 2002, 2011) whose mission is dedicated to modeling the static Earth's gravitational field to high resolution (maximum harmonic degree, 250). The GOCE gradiometer consists of three pairs of three-degree-of-freedom accelerometers that sense all nine gradient tensor elements, although only the diagonal and a third of the off-diagonal elements achieve the highest performance due to limitations in calibration. The spectral band of the measurements is 0.005–0.1 Hz and the noise floor is about $10-20\times10^{-12}\ \mathrm{s}^{-2}/\sqrt{\mathrm{Hz}}$.

Both CHAMP and GRACE have yielded global gravitational models by utilizing traditional satellite tracking methods and incorporating the range rate appropriately as a tracking observation (time-wise approach). However, the immediate application of eqn [133] suggests that gravitational differences can be determined *in situ* and used to determine a model for the global field directly. This is classified as the space-wise approach. In fact, if the orbits are known with sufficient accuracy (from kinematic orbit determination, e.g., by GPS), this procedure utilizes a linear relationship between observations and unknown harmonic coefficients:

$$\left.\frac{\mathrm{d}^2\rho}{\mathrm{d}t^2}\right|_{x_1,x_2} = \sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n} v_{nm}\boldsymbol{e}_\rho^{\mathrm{T}}\left(\boldsymbol{\nabla}U_{nm}(\theta,\lambda,r)|_{x_2} - \boldsymbol{\nabla}U_{nm}(\theta,\lambda,r)|_{x_1}\right) + \delta a + \delta c \qquad [134]$$

where $U_{nm}(\theta,\lambda,r) = (R/r)^{n+1}\overline{Y}_{nm}(\theta,\lambda)$ and δa and δc are the last two terms in eqn [133]. Given the latter and a set of line-of-sight accelerations, a theoretically straightforward linear least-squares adjustment solves for the coefficients. A similar procedure can be used for gradients observed on an orbiting satellite, such as GOCE:

$$\left.\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{b}}\right|_{\boldsymbol{x}} = \sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n} v_{nm}\boldsymbol{\nabla}\boldsymbol{\nabla}^{\mathrm{T}}U_{nm}(\theta,\lambda,r)|_{x} + \boldsymbol{\psi} \qquad [135]$$

where $\boldsymbol{\psi}$ comprises the rotational acceleration terms in eqn [132].

Recently, a rather different theory has been considered by several investigators to model the gravitational field from satellite-to-satellite tracking observations. The method, first proposed by Wolff (1969), makes use of yet another fundamental law: the law of conservation of energy. Simply, the range rate between two satellites implies an along-track velocity difference or a difference in kinetic energy. Observing this difference leads directly, by the conservation law, to the difference in potential energy, that is, the gravitational potential and other potential energies associated with action forces and Earth's rotation. Neglecting the latter two, conservation of energy implies

$$V = \frac{1}{2}\left|\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}_1\right|^2 - E_0 \qquad [136]$$

where E_0 is a constant. Taking the along-track differential, we have approximately

$$V(\boldsymbol{x}_2) - V(\boldsymbol{x}_1) \approx \delta V = \left|\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}_1\right| \frac{\mathrm{d}}{\mathrm{d}t}\rho \qquad [137]$$

where $|\delta(\mathrm{d}\boldsymbol{x}_1/\mathrm{d}t)| \approx |\mathrm{d}\boldsymbol{x}_2/\mathrm{d}t - \mathrm{d}\boldsymbol{x}_1/\mathrm{d}t| \approx \mathrm{d}\rho/\mathrm{d}t$. This very rough conceptual relationship between the potential difference and the range rate applies to two satellites closely following each other in similar orbits. The precise formulation is given by Jekeli (1999) and holds for any pair of satellites, not just two low orbiters. Mapping range rates between two polar orbiting satellites (such as GRACE) yields a global distribution of potential difference observations related again linearly to a set of harmonic coefficients:

$$V(\boldsymbol{x}_2) - V(\boldsymbol{x}_1) = \sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n} v_{nm}\left(U_{nm}(\theta,\lambda,r)|_{x_2} - U_{nm}(\theta,\lambda,r)|_{x_1}\right) \qquad [138]$$

The energy-based model holds for any two vehicles in motion and equipped with the appropriate ranging and accelerometry instrumentation. For example, the energy conservation principle could also be used to determine geopotential differences between an aircraft and a satellite, such as a GPS satellite (at which location the geopotential is known quite well). The aircraft would require only a GPS receiver (to do the ranging) and a set of accelerometers (and gyros for orientation) to measure the action forces (the same system components as in airborne accelerometry discussed earlier). Resulting potential differences could be used directly to model the geoid undulation differences using Bruns' equation [3].

The *in situ* measurements of line-of-site acceleration or of more local gradients would need to be reduced from the satellite orbit to a well-defined surface (such as a sphere) in order to serve as boundary values in a solution to a BVP. However, the model for the field is already in place, rooted in potential theory (truncated series solution to Laplace's equation), and one may think of the problem more in terms of fitting a three-dimensional model to a discrete set of observations. This operational approach can readily, at least conceptually, be expanded to include observations from many different satellite systems, even airborne and ground-based observations.

3.02.6.2 Models

Global spherical harmonic models have been derived since the first satellite launches in the 1960s using a variety of methods and combinations of terrestrial data and satellite tracking and *in situ* data. The recent gravity satellite missions have generated a plethora of such models, which are collected and organized by the International Center for Earth Gravitational Models (ICEGM (http://icgem.gfz-potsdam.de/ICGEM/)) of the International Association of Geodesy (IAG). The spherical harmonic models come in two varieties, either based strictly on satellite data, in which case, their resolution is limited as discussed earlier or combined with terrestrial data, including satellite altimetry over the oceans, for vastly higher resolution. The responsible centers for collecting, distributing, and archiving the raw satellite data make these (satellite) data and the derived global models available for general scientific investigations.

The standard solution to the BVP (**Section 3.02.4**), using terrestrial gravimetry in the form of gravity anomalies, now routinely incorporates a longer-wavelength global model derived from satellite data (such as GRACE or GOCE). The solution is either regional using Green's function approach, Stokes' integral (eqn [72]), or global, based on the harmonic analysis of surface data, either using the integrals [75] or solving a linear system of equations (eqn [71] truncated to finite degree) to obtain the higher-degree coefficients, δC_{nm}. The integrals must be evaluated using quadratures, and very fast numerical techniques have been developed when the data occupy a regular grid of coordinates on the sphere or ellipsoid (Rapp and Pavlis, 1990). Similar algorithms enable the fast solution of the linear system of eqn [71].

For the global harmonic analysis, the number of coefficients, $(n_{\max}+1)^2$, must not be greater than implied by the resolution of the data. A general, conservative rule of thumb for the relationship between the maximum resolution (half-wavelength), $\Delta\theta$, in angular degrees on the unit sphere, and the maximum degree of a truncated spherical harmonic series is

$$\Delta\theta = \frac{180^\circ}{n_{\max}} \qquad [139]$$

Thus, data on a $1\times1^\circ$ angular grid of latitudes and longitudes would imply $n_{\max}=180$. The number of data (64 800, in this

case) is amply larger than the number of coefficients (32 761). This majority suggests a least-squares adjustment of the coefficients to the data, in either method, especially because the data have errors (Rapp, 1969). As $n_{\max}$ increases, a rigorous, optimal adjustment usually is feasible, for a given computational capability, only under restrictive assumptions on the correlations among the errors in the data. Also, the obvious should be noted that the accuracy of the global model in any area depends on the quality of the data in that area. Furthermore, considering that a measurement implicitly contains all harmonics (up to the level of measurement error), the estimation of a finite number of harmonics from boundary data on a given grid is corrupted by those harmonics that are in the data but are not specifically estimated. This phenomenon is called aliasing in spectral analysis and can be mitigated by appropriate filtering of the data (Jekeli, 1996).

The currently best known and arguably the overall most accurate model is EGM08 (Earth Gravitational Model 2008) complete to degree and order $n_{\max}=2160$ (5′ resolution) with additional higher-degree harmonics up to 2190 (Pavlis et al., 2012). This model updates the previous EGM96 (Lemoine et al., 1998) for the WGS84 of the US Department of Defense by including improved gravity anomalies over the oceans derived from satellite altimetry, additional land data sets, and airborne gravimetry and the GRACE gravitational field model to maximum degree and order, $n_{\max}=180$. This latter model was used with its formal covariance error measure and replaced the previously used satellite tracking data that defined the lower-degree harmonics of EGM96. In constructing combination solutions of this type, great effort is expended to ensure the proper weighting of all observations in order to extract the most appropriate information from the diverse data, pertaining to different parts of the spatial gravitational spectrum. It is beyond the present scope to delve into the detailed numerical methodology of combination methods; useful starting points are the documents for EGM96 and EGM08, cited earlier.

Stokes' integral is used in practice today only to take advantage of local or regional data with higher resolution than available from the global models. Even though the integral is a global integral, it can be truncated to a neighborhood of the computation point since Stokes' kernel attenuates rapidly with distance from its origin. Moreover, the corresponding truncation error may be reduced if the boundary values exclude the longer-wavelength features of the field. The latter constitute an adequate representation of the remote zone contribution and can be included separately as follows. Let $\Delta g^{(n_{\max})}$ denote the gravity anomaly implied by a spherical harmonic model, such as given by eqn [71], truncated to degree, $n_{\max}$. From the orthogonality of spherical harmonics (eqn [52]), it is easy to show that

$$\begin{aligned} T^{(n_{\max})}(\theta,\lambda,r) &= \frac{GM}{R}\sum_{n=2}^{n_{\max}}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+1}\delta C_{nm}\overline{Y}_{nm}(\theta,\lambda) \\ &= \frac{R}{4\pi}\iint_{\sigma}\Delta g^{(n_{\max})}\left(\theta',\lambda',R\right)S(\psi,r)\,\mathrm{d}\sigma \end{aligned} \qquad [140]$$

Thus, given a spherical harmonic model $\{\delta C_{nm}|\ 2\le n\le n_{\max},\ -n\le m\le n\}$, one first removes the model in terms of the gravity anomaly and then restores it in terms of the disturbing potential, changing Stokes' formula [72] to

$$\begin{aligned} T(\theta,\lambda,r) = \frac{R}{4\pi}\iint_{\sigma}\left(\Delta g\left(\theta',\lambda',R\right)-\Delta g^{(n_{\max})}\left(\theta',\lambda',R\right)\right)S(\psi,r)\,\mathrm{d}\sigma \\ + T^{(n_{\max})}(\theta,\lambda,r) \end{aligned} \qquad [141]$$

In theory, if $\Delta g^{(n_{\max})}$ has no errors, then the residual $\Delta g-\Delta g^{(n_{\max})}$ excludes all harmonics of degree $n\le n_{\max}$, and orthogonality would also allow the exclusion of these harmonics from S. Once the integration is limited to a neighborhood of (θ,λ,R), as it must be in practice, there are a number of ways to modify the kernel so as to minimize the resulting truncation error (Sjöberg, 1991, and references therein); for a review, see also Featherstone (2013). The removal and restoration of a global model, however, is the key aspect in all these methods.

In practical applications, the boundary values are on the geoid, being the surface that satisfies the boundary condition of the Robin or Neumann BVP (**Section 3.02.3.2**); that is, we require the normal derivative on the boundary, and measured gravity is indeed the derivative of the potential along the perpendicular to the geoid. The integral in eqn [141] thus approximates the geoid by a sphere. Furthermore, it is assumed that no masses exist external to the geoid. Part of the reduction to the geoid of data measured on the Earth's surface involves redistributing the topographic masses on or below the geoid. This redistribution is undone outside the solution to the BVP (i.e., Stokes' integral) in order to regain the disturbing potential for the actual Earth. Conceptually, we may write

$$T_P=\frac{R}{4\pi}\iint_{\sigma}\left(\Delta g-\Delta g^{(n_{\max})}-\delta c\right)_{P'}S_{P,P'}\ \ \mathrm{d}\sigma+T_P^{(n_{\max})}+\delta T_P \qquad [142]$$

where δc is the gravity reduction that brings the gravity anomaly to a geoid with no external masses and δT_P is the effect (called indirect effect) on the disturbing potential due to reversing this reduction. This formula holds for T anywhere on or above the geoid and thus can also be used to determine the geoid undulation according to Bruns' formula [3].

3.02.7 The Geoid and Heights

The traditional reference surface, or datum, for heights is the geoid (**Section 3.02.1.3**). A point at mean sea level usually serves as starting point (datum origin), and this defines the datum for vertical control over a region or country. The datum (or regional geoid) is the level continuation of the reference surface under the continents; and the determination of gravity potential differences from the initial point to other points on the Earth's surface, obtained by leveling and gravity measurements, yields heights with respect to that reference (or in that datum). The gravity potential difference, known as the geopotential number, at a point, P, relative to the datum origin, $\overline{P}_0$, is given by (since gravity is the negative vertical derivative of the gravity potential)

$$C_P=W_0-W_P=\int_{\overline{P}_0}^{P}g\,\mathrm{d}n \qquad [143]$$

where g is gravity magnitude, dn is a leveling increment along the vertical direction, and W_0 is the gravity potential at $\overline{P}_0$. By the conservative nature of the gravity potential (eqn [24]), whatever path is taken for the integral yields a unique geopotential number for P. From these potential differences, one can define various types of height, for example, the orthometric height (Figure 3):

$$H_P = \frac{C_P}{\overline{g}_P} \qquad [144]$$

where

$$\overline{g}_P = \frac{1}{H_P}\int_{P_0}^{P} g\,\mathrm{d}H \qquad [145]$$

is the average value of gravity along the plumb line from the geoid at P_0 to P. Other height systems are also in use, such as the normal and dynamic heights:

$$H_P^N = \frac{C_P}{\overline{\gamma}_P}, \quad \overline{\gamma}_P = \frac{1}{H_P^N}\int_{Q_0}^{Q} \gamma\,\mathrm{d}H^N \qquad [146]$$

$$H_P^{\mathrm{dyn}} = \frac{C_P}{\gamma_0}, \quad \gamma_0 = \text{constant} \qquad [147]$$

where reference is made to Hofmann-Wellenhof and Moritz (2005, **Chapter 1.05**), Jekeli (2000b), and NGS (1986) for additional details.

For a particular height datum, there is theoretically only one datum surface (the geoid). But access to this surface is far from straightforward at points other than at the defined datum origin. If $\overline{P}_0$ is defined at mean sea level, other points at mean sea level are not on the same level surface, since mean sea level, in fact, is not level. Erroneously assuming that mean sea level is an equipotential surface can cause significant distortions in the vertical control network of larger regions, as much as several decimeters. This was the case, for example, for the National Geodetic Vertical Datum of 1929 in the United States for which 26 mean sea level points on the East and West Coasts were assumed to lie on the same level surface. Accessibility to the geoid (once defined) at any point is achieved either with precise leveling and gravity, according to eqns [143] and [144], or with precise geometric vertical positioning and knowledge of the gravity potential, as described later in this section. Indeed, as models for the geopotential improve, traditional vertical data may be replaced with a geopotential-based datum that completely eliminates the need for leveling. As noted in **Section 3.02.1**, Canada (in 2013) and the United States (to be implemented in the early 2020s) have embarked on this new paradigm for vertical control.

Geometric vertical positioning, today, is obtained very accurately (cm accuracy or even better) with differential GPS. Suppose that an accurate gravity potential model is also available in the same coordinate system as used for GPS. Then, determining the GPS position at $\overline{P}_0$ allows the evaluation of the gravity potential, W_0, of the datum. From a practical standpoint, one may choose many points near sea level and define an average value of W_0 as the datum or geoid potential. Access to the geoid at any other point, P, or equivalently, determining the orthometric height, H_P, can be done by first determining the ellipsoidal height, h, from GPS. Then, as shown in Figure 3,

$$H_P = h_P - N \qquad [148]$$

where, with $T = W - U$ evaluated on the geoid, Bruns' extended equation (**Section 3.02.1**) yields

$$N = T/\gamma - (W_0 - U_0)/\gamma \qquad [149]$$

where U_0 is the normal gravity potential of the normal ellipsoid. Here, it is noted that models for T (**Section 3.02.6.2**) generally exclude a constant term, which implies that T/γ is the height of a global geoid with respect to a best-fitting ellipsoid. For a predefined ellipsoid (given U_0), and a predefined datum point, $\overline{P}_0$ (or geopotential value, W_0), the second term in eqn [149] is determined independently of the geopotential model.

In a sense, determining N requires N in order to locate the point of computation of T on the geoid. However, only an approximate height, H, suffices to reduce the value of T onto the geoid, where because of the vertical gradient of the order of 5×10^{-4} m s^{-2}, a height error of 20 m leads to an error of 10^{-2}m^2 s^{-2} in T or just 1 mm in the geoid undulation. It should be noted that a model for the disturbing potential as a series of spherical harmonics, for example, derived from satellite observations, satisfies Laplace's equation and, therefore, does not give the correct disturbing potential at the geoid (if it lies below the Earth's surface where Laplace's equation does not hold). Details for the proper evaluation of the geoid undulation from spherical harmonic series of the geopotential are given by Lemoine et al. (1998, **Chapter 1.13**).

The ability to derive orthometric heights (or other geopotential-related heights) from GPS has enormous economical advantage over the alternative and laborious leveling procedure. This has put great emphasis on obtaining an accurate geoid undulation model for land areas. **Section 3.02.6** briefly outlined the essential methods to determine the geopotential from a combination of spherical harmonics and an integral of local gravity anomalies. Bruns' equation then yields the geoid undulation; and as noted earlier, when dealing with a height datum or a regional geoid, the constant $N_0 = -(W_0 - U_0)/\gamma$ requires careful attention (Lemoine et al., 1998, **Chapter 1.13**). It can be determined by comparing the geoid undulation computed according to a model that excludes this term (such as eqn [141]) with at least one geoid undulation (usually many) determined from leveling and GPS, according to eqn [148]; for an example, see Jekeli et al. (2012). Vertical control and the choice of height datum are specific to each country or continent, where a local mean sea level was the adopted datum origin. Thus, height data around the world are 'local geoids' or 'regional geoids' that have significant level differences between them. Investigations and efforts have been under way for many decades to define a global vertical datum; however, an internationally implemented geoid is still in the future, as global geopotential models become more accurate and, perhaps more crucially, as unavoidable differences in mean sea level between conventional datums become secondary to absolute global vertical control. The current convention within the IERS (Petit and Luzum, 2010) is to define the global geoid by the potential value, $W_0 = 62636856.0 \pm 0.5$ m^2 s^{-2}.

On the oceans, the situation is somewhat less complicated. Oceanographers who compute sea surface topography from satellite altimetry on the basis of eqn [148] depend critically

on an accurate geoid undulation or equivalently on an accurate model of T. However, no reduction of the disturbing potential from mean sea level to the geoid is necessary; the deviation between the two being at most 2 m, and if neglected, causes an error in geoid undulation of less than 0.1 mm. Thus, a high-degree spherical harmonic model of T is entirely appropriate. Furthermore, it is reasonable to require that the constant, $W_0 - U_0$, vanishes over the oceans. That is, one may choose the geoid such that it best fits global mean sea level and choose an ellipsoid that best fits this geoid. It means that the global average value of the geoid undulation should be zero (according to eqn [149]). The latter can be achieved with satellite altimetry and oceanographic models of sea surface topography (Bursa et al., 1997).

Several interesting and important distinctions should be made in regard to the tidal effects on the geoid. The sun and the moon generate an appreciable gravitational potential near the Earth (the other planets may be neglected). In an Earth-fixed coordinate system, this extraterrestrial potential varies in time with different periods due to the relative motions of the moon and sun and because of Earth's rotation (Torge and Müller, 2012, p. 88). There is also a constant part, the permanent tidal potential, representing the average over time. It is not zero because the Earth–sun–moon system is approximately coplanar. For each extraterrestrial body with mass, M_B, and distance, r_B, from the Earth's center, this permanent part is given by

$$V_c^B(\theta, r) = \frac{3}{4} G M_B \frac{r^2}{r_B^2} \left(3\cos^2\theta - 1\right)\left(\frac{1}{2}\sin^2\varepsilon - \frac{1}{3}\right) \qquad [150]$$

where ε is the angle of the ecliptic relative to the equator ($\varepsilon \approx 23°44'$). Using nominal parameter values for the sun and the moon, we obtain at mean earth radius, $R = 6371$ km,

$$V_c^{s+m}(\theta, R) = -0.97\left(3\cos^2\theta - 1\right)\mathrm{m}^2\mathrm{s}^{-2} \qquad [151]$$

The gravitational potential from the sun and moon also deforms the quasi-elastic Earth's masses with the same periods and similarly includes a constant part. These mass displacements (both ocean and solid Earth) give rise to an additional indirect change in potential, the tidal deformation potential (there are also secondary indirect effects due to loading of the ocean on the solid Earth, which can be neglected in this discussion). The indirect effect is modeled as a fraction of the direct effect (Lambeck, 1988, p. 254), so that the permanent part of the tidal potential including the indirect effect is given by

$$\overline{V}_c^{s+m}(\theta, R) = (1 + k_2) V_c^{s+m}(\theta, R) \qquad [152]$$

where $k_2 = 0.29$ is Love's number (an empirical number based on observation). This is also called the mean tidal potential.

The mean tidal potential is inherent in all our terrestrial observations (the boundary values) and cannot be averaged away in time; yet, the solutions to the BVP assume no external masses. Therefore, in principle, the effect of the tidal potential including its mean, or permanent, part should be removed from the observations prior to applying the BVP solutions. On the other hand, the permanent indirect effect is not that well modeled and arguably should not be removed; after all, it contributes to the Earth's shape as it actually is in the mean. Three types of tidal systems have been defined to distinguish between these corrections. A *mean* quantity refers to the quantity that retains the mean tidal potential (but time-varying parts are removed); a *nontidal* quantity implies that all tidal effects (time-varying, permanent, direct, and indirect effects) have been removed computationally; and the *zero-tide* quantity excludes all time-varying parts and the permanent direct effect, but it retains the indirect permanent effect.

If the geoid (an equipotential surface) is defined solely by its potential, W_0, then a change in the potential due to the tidal potential, V^{tide} (time-varying and constant parts and direct and indirect effects), implies that the W_0-equipotential surface has been displaced. The geoid is now a different surface with the same W_0. This displacement is equivalent to a change in geoid undulation, $\delta N = V^{\text{tide}}/\gamma$, with respect to some predefined ellipsoid. The permanent tidal effect (direct and indirect) on the geoid is given by

$$\delta\overline{N}(\theta) = -0.099(1 + k_2)\left(3\cos^2\theta - 1\right)\mathrm{m} \qquad [153]$$

If N represents the instantaneous geoid undulation, then the geoid undulation without any tidal effects, that is, the nontidal geoid undulation, is given by

$$N_{\text{nt}} = N - \delta N \qquad [154]$$

The mean geoid undulation is defined as the geoid undulation with all but the mean tidal effects removed:

$$\overline{N} = N - \left(\delta N - \delta\overline{N}\right) \qquad [155]$$

It is the geoid undulation observed, for example, using satellite altimetry, averaged over time (considering only tide-induced variations). We omit here a discussion of secular changes in the geoid due to long-term deformation of the Earth due, in particular, to postglacial isostatic adjustments in the higher north and south latitudes. The zero-tide geoid undulation retains the permanent indirect effect, but no other tidal effects:

$$N_z = N - \left(\delta N + 0.099 k_2\left(3\cos^2\theta - 1\right)\right)\mathrm{m} \qquad [156]$$

The difference between the mean and zero-tide geoids is, therefore, the permanent component of the direct tidal potential. We note that, in principle, each of the geoids defined earlier has the same potential value, W_0, in its own field. That is, with each correction, we define a new gravity field and the corresponding geoid undulation defines the equipotential surface in that field with potential value given by W_0. This is fundamentally different than what happens in the case when the geoid is defined as a vertical datum using a specified datum origin point. In this case, one needs to consider also the vertical displacement of the datum point due to the tidal deformation of the Earth's surface. The potential of the datum then changes because of the direct tidal potential, the indirect effect due to mass changes, and the indirect effect due to the vertical displacement of the datum (for additional details, see Jekeli, 2000b).

References

Bursa M, Radej K, Sima Z, True S, and Vatrt V (1997) Determination of the geopotential scale factor from TOPEX/Poseidon satellite altimetry. *Studia Geophysica et Geodaetica* 41: 203–216.

Cappelari JO, Velez CE, and Fuchs AJ (eds.) (1976) *Mathematical Theory of the Goddard Trajectory Determination System*. Greenbelt, MD: Goddard Space Flight Center, GSFC Document X-582-76-77.

Chao BF (2005) On inversion for mass distribution form global (time-variable) gravity field. *Journal of Geodynamics* 29: 223–230.

Courant R and Hilbert D (1962) *Methods of Mathematical Physics*, vol. 2. New York: Wiley.

Cushing JT (1975) *Applied Analytical Mathematics for Physical Scientists*. New York: Wiley.

Dziewonsky AD and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297.

Featherstone WE (2013) Deterministic, stochastic, hybrid and band-limited modifications of Hotine's integral. *Journal of Geodesy* 87(5): 487–500.

Fei ZL and Sideris MG (2000) A new method for computing the ellipsoidal correction for Stokes's formula. *Journal of Geodesy* 74(2): 223–231.

Goodkind JM (1999) The superconducting gravimeter. *Review of Scientific Instruments* 70: 4131–4152.

Gumert WR (1998) An historical review of airborne gravimetry. *The Leading Edge* 17(1): 113–116.

Günter NM (1967) *Potential Theory and its Applications to Basic Problems of Mathematical Physics*. New York: Frederick Ungar Publishing Co.

Hammer S (1983) Airborne gravity is here. *Geophysics* 48(2): 213–223.

Heiskanen WA and Moritz H (1967) *Physical Geodesy*. San Francisco, CA: Freeman and Co.

Hobson EW (1965) *The Theory of Spherical and Ellipsoidal Harmonics*. New York: Chelsea Publishing Co.

Hofmann-Wellenhof B and Moritz H (2005) *Physical Geodesy*. Berlin: Springer.

Hotine M (1969) *Mathematical Geodesy*. Washington, D.C.: U.S. Department of Commerce.

Huang J and Véronneau M (2013) Canadian gravimetric geoid model 2010. *Journal of Geodesy* 87: 771–790.

Jekeli C (1988) The exact transformation between ellipsoidal and spherical harmonic expansions. *Manuscripta Geodaetica* 14: 106–113.

Jekeli C (1996) Spherical harmonic analysis, aliasing, and filtering. *Journal of Geodesy* 70: 214–223.

Jekeli C (1999) The determination of gravitational potential differences from satellite-to-satellite tracking. *Celestial Mechanics and Dynamical Astronomy* 75(2): 85–100.

Jekeli C (2000a) *Inertial Navigation Systems with Geodetic Applications*. Berlin: Walter deGruyter.

Jekeli C (2000) Heights, the geopotential, and vertical datums. Report no. 459. Columbus, OH: Ohio State University, http://www.geology.osu.edu/jekeli.1/OSUReports/reports/report_459.pdf.

Jekeli C (2004) High-resolution gravity mapping: the next generation of sensors. In: Sparks RSJ (ed.) *The State of the Planet: Frontiers and Challenges in Geophysics*. *Geophysical Monograph 150*, vol. 19, pp. 135–146. IUGG.

Jekeli C, Yang HJ, and Kwon JH (2012) The offset of the South Korean vertical datum from a global geoid. *KSCE Journal of Civil Engineering* 16(5): 816–821.

Kaula WM (1966) *Theory of Satellite Geodesy*. London: Blaisdell Publishing Co.

Kellogg OD (1953) *Foundations of Potential Theory*. New York: Dover Publishing.

Kenyon S and Forsberg R (2000) Arctic gravity project. *SEG Technical Program Expanded Abstracts* 19: 410–413.

Kwon JH and Jekeli C (2001) A new approach for airborne vector gravimetry using GPS/INS. *Journal of Geodesy* 74: 690–700.

Lambeck K (1988) *Geophysical Geodesy*. Oxford: Clarendon Press.

Lane RJL (ed.) (2004) *Airborne Gravity 2004 – Abstracts from the ASEG-PESA Airborne Gravity 2004 Workshop*, Geoscience Australia Record 2004/18.

Lane RJL (ed.) (2010) *Airborne Gravity 2010 – Abstracts from the ASEG-PESA Airborne Gravity 2010 Workshop*, Published jointly by Geoscience Australia and the Geological Survey of New South Wales, Geoscience Australia Record 2010/23 and GSNSW File GS2010/0457.

Lemoine FG, Kenyon SC, Factor JK, et al. (1998) *The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96, NASA Technical Paper NASA/TP-1998-206861*. Greenbelt: Goddard Space Flight Center.

Martin JL (1988) *General Relativity: A Guide to its Consequences for Gravity and Cosmology*. New York: Wiley.

Mohr PJ, Taylor BN, and Newell DB (2012) CODATA recommended values of the fundamental physical constants: 2010. *Reviews of Modern Physics* 84: 1527–1605.

Molodensky MS, Eremeev VG, and Yurkina MI (1962) *Methods for Study of the External Gravitational Field and Figure of the Earth*. Translation from Russian; Israel Program for Scientific Translations, Jerusalem.

Moritz H (1980) *Advanced Physical Geodesy*. Karlsruhe: Wichmann, Reprint 2001 by Civil and Environmental Engineering and Geodetic Science, Ohio State University, Columbus, OH.

Moritz H (2000) Geodetic Reference System 1980. *Journal of Geodesy* 74(1): 128–133.

Morse PM and Feshbach H (1953) *Methods of Theoretical Physics, Parts I and II*. New York: McGraw-Hill Book Co.

Müller C (1966) *Spherical harmonics*. *Lecture Notes in Mathematics*. Berlin: Springer.

Nettleton LL (1976) *Gravity and Magnetics in Oil Prospecting*. New York: McGraw-Hill Book Co.

NGS (1986) *Geodetic Glossary*. Rockville, MD: NOAA/NOS, National Geodetic Information Center, Publication of the National Geodetic Survey.

NGS (2012) *National Height Modernization Strategic Plan*. Draft Document Released to Public, November 2012, National Geodetic Survey, NOAA, Silver Spring, MD. http://www.ngs.noaa.gov/web/news/HMODPlan_PublicReview_Nov2012.pdf.

Niebauer T, Sasagawa G, Faller J, Hilt R, and Klopping F (1995) A new generation of absolute gravimeters. *Metrologia* 32(3): 159–180.

NIMA (2000) The Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems. National Imagery and Mapping Agency Technical Report TR-8350.2, 3rd edn. Bethesda, MD.

Olesen AV and Forsberg R (2007) Airborne scalar gravimetry for regional gravity field mapping and geoid determination. In: *Proceedings of the Symposium of the International Gravity Field Service, 28 August–1 September 2006, Istanbul, Turkey*, Harita Dergisi (Journal of Mapping), vol. 18, pp. 277–282, General Command of Mapping, Ankara, Turkey.

Pavlis NK, Holmes SA, Kenyon SC, and Factor JF (2012) The development and evaluation of Earth Gravitational Model (EGM2008). *Journal of Geophysical Research* 117. http://dx.doi.org/10.1029/2011JB008916.

Pavlis DE, Moore D, Luo S, McCarthy JJ, and Luthcke SB (1999) *GEODYN Operations Manual*. vol. 5: Greenbelt, MD: Raytheon ITSS.

Petit G and Luzum B (2010) IERS Conventions (2010). IERS Technical Note No. 36, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main.

Petrovic S (1996) Determination of the potential of homogeneous polyhedral bodies using line integrals. *Journal of Geodesy* 71: 44–52.

Petrovskaya MS (1979) Upward and downward continuations in the geopotential determination. *Bulletin Géodésique* 53: 259–271.

Plag HP and Pearlman M (eds.) (2009) *Global Geodetic Observing System, Meeting the Requirements of a Global Society on a Changing Planet in 2020*. Dordrecht: Springer.

Rapp RH (1969) Analytical and numerical differences between two methods for the combination of gravimeter and satellite data. *Bollettino di Geofisica Teorica ed Applicata* 11(41–42): 108–118.

Rapp RH and Pavlis NK (1990) The development and analysis of geopotential coefficient models to spherical harmonic degree 360. *Journal of Geophysical Research* 95(B13): 21885–21911.

Reigber C, Balmino G, Schwintzer P, et al. (2002) A high quality global gravity field model from CHAMP GPS tracking data and accelerometry (EIGEN-1S). *Geophysical Research Letters* 29(14). http://dx.doi.org/10.1029/2002GL015064.

Rummel R, Balmino G, Johannessen J, Visser P, and Woodworth P (2002) Dedicated gravity field missions – principles and aims. *Journal of Geodynamics* 33: 3–20.

Rummel R, Yi W, and Stummer C (2011) GOCE gravitational gradiometry. *Journal of Geodesy* 85: 777–790.

Sandwell DT and Smith WHF (1996) Marine gravity anomaly form GEOSAT and ERS-1 satellite altimetry. *Journal of Geophysical Research* 102(B5): 10039–10054.

Sansò F and Rummel R (eds.) (1997) *Geodetic boundary value problems in view of the one centimeter geoid*. In: *Lecture Notes in Earth Sciences*, vol. 65. Berlin: Springer.

Sebera J, Bouman J, and Bosch W (2012) On computing ellipsoidal harmonics using Jekeli's renormalization. *Journal of Geodesy* 86(9): 713–726.

Seeber G (1993) *Satellite Geodesy, Foundations, Methods, and Applications*. Berlin: W. de Gruyter.

Sjöberg LE (1991) Refined least squares modifications of Stokes' formula. *Manuscripta Geodaetica* 16: 367–375.

Studinger M, Bell R, and Frearson N (2008) Comparison of AIRGrav and GT-1A airborne gravimeters for research applications. *Geophysics* 73(6): 151–161.

Tapley BD (1973) Statistical orbit determination theory. In: Tapley BD and Szebehely V (eds.) *Recent Advances in Dynamical Astronomy*, pp. 396–425. Dordtecht: D. Reidel Publishing Co.

Tapley BD, Bettadpur S, Ries JC, Thompson PF, and Watkins M (2004) GRACE measurements of mass variability in the Earth system. *Science* 305(5683): 503–505.

Tapley BD, Bettadpur S, Watkins M, and Reigber C (2004) The Gravity Recovery and Climate Experiment, mission overview and early results. *Geophysical Research Letters* 31(9). http://dx.doi.org/10.1029/2004GL019920.

Tapley BD, Ries J, Bettadpur S, et al. (2005) GGM02 – An improved Earth gravity field model from GRACE. *Journal of Geodesy*. http://dx.doi.org/10.1007/s00190-005-0480-z.
Telford WM, Geldart LP, and Sheriff RE (1990) *Applied Geophysics*, 2nd edn. Cambridge, U.K.: Cambridge University Press.
Torge W (1989) *Gravimetry*. Berlin: Walter de Gruyter.
Torge W and Müller J (2012) *Geodesy*, 4th edn Berlin: Walter de Gruyter.
Tsoulis D and Petrovic S (2001) On the singularities of the gravity field of a homogeneous polyhedral body. *Geophysics* 66(2): 535–539.
Wolff M (1969) Direct measurement of the earth's gravitational potential using a satellite pair. *Journal of Geophysical Research* 74: 5295–5300.
Yu J, Jekeli C, and Zhu M (2002) The analytical solutions of the Dirichlet and Neumann boundary value problems with ellipsoidal boundary. *Journal of Geodesy* 76(11–12): 653–667.
Zumberge MA, Rinker RL, and Faller JE (1982) A portable apparatus for absolute measurements of the Earth's gravity. *Metrologia* 18(3): 145–152.

This page intentionally left blank

3.03 Gravimetric Methods – Absolute and Relative Gravity Meter: Instruments Concepts and Implementation

T Niebauer, Micro-g LaCoste, LLC, Lafayette, CO, USA

© 2015 Elsevier B.V. All rights reserved.

3.03.1	**Absolute and Relative Gravity Meters**	37
3.03.1.1	Definition of Absolute and Relative Gravity and Their Practical Uses	37
3.03.1.2	Definition of Precision and Accuracy	38
3.03.1.2.1	Precision	39
3.03.1.2.2	Repeatability	41
3.03.1.2.3	Accuracy	42
3.03.2	**Gravity Meters**	43
3.03.2.1	Absolute Gravity Meters	44
3.03.2.1.1	Pendulum absolute gravity meters	44
3.03.2.1.2	Free-fall absolute gravity meters	45
3.03.2.1.3	Commercial absolute gravity meters	50
3.03.2.2	Relative Gravity Meters	51
3.03.2.2.1	The L&R zero-length spring suspension	51
3.03.2.2.2	Quartz vertical spring gravity meters	53
3.03.2.2.3	The superconducting gravity meter	54
3.03.2.3	Gravity Meters Applications	54
References		56

3.03.1 Absolute and Relative Gravity Meters

3.03.1.1 Definition of Absolute and Relative Gravity and Their Practical Uses

Terrestrial gravity measurements are made to determine the magnitude of the acceleration of gravity due to the Earth along the direction of a locally freely falling body. The local magnitude of gravity, denoted by the familiar symbol '*g*,' is referred to as 'little *g*.' In most physics textbooks, *g* is assumed to be a constant value of $9.8\ \mathrm{m\,s^{-2}}$ (often rounded to $10\ \mathrm{m\,s^{-2}}$). There is typically very little discussion in introductory physics books about gravity and usually no mention that it actually varies from point to point on the Earth or that it can be of practical interest to measure its value. The two most commonly asked questions by college-educated persons are (1) is gravity not the same everywhere and (2) why would anyone want to measure gravity? It seems to be a well-kept secret that the measurement of gravity on the Earth's surface can provide a wealth of information about the shape of the Earth, the position of the continents, and the water and mineral contents of the ground underneath our feet.

Measurement of gravity has a long and rich history closely connected to the measurement of time. The first accurate pendulums were used to measure time and also to measure gravity. Currently, the most accurate measurements of gravity are still linked to very precise time measurements of the trajectory for a free-falling object. Gravity measurements have been used to answer questions about the Earth's gravity field since the time of Galileo (1564–1642) who purportedly dropped stones from the Leaning Tower of Pisa to test Aristotle's theory that heavy things fall faster than light objects. Gravity is measured on the surface of the Earth with land-based gravity meters on the oceans and from airplanes using dynamic gravity meters. Since March 2002, the Gravity Recovery and Climate Experiment, a joint mission of NASA and the German Aerospace Center, is continuously measuring the gravity field of the Earth from space. The mission uses a microwave ranging system to accurately measure changes in the speed and distance between two identical spacecraft flying in a polar orbit about 220 km apart, 500 km above Earth. These gravity measurements are now precise enough to start answering very important time-lapse questions about global warming and sea-level rise (Jacob et al., 2012). Gravity is also used to answer more localized questions about the amount of water, oil, and minerals that are buried deep within the ground. Other space measurements like NASA's Gravity Probe B (Everitt et al., 2011) test Einstein's theory of relativity.

Real-world gravity measurements historically have been made using the CGS unit called a gal, which is named after Galileo (or Gal) who is often considered the father of gravity. In these units, the Earth's field is given as 980 Gal (or ~1000 Gal), where $1\ \mathrm{Gal} = 1\ \mathrm{cm\,s^{-2}}$. A calibrated measurement of the acceleration of gravity given in absolute units of distance and time ($\mathrm{m\,s^{-2}}$ in SI units) is referred to as an 'absolute measurement of gravity.'

Terrestrial gravity measurements are higher at the poles than at the equator by several parts in 1000 (several Gal) due to a combination of a decreasing distance to the center of the Earth and less centrifugal acceleration at the poles. Thus, gravity is obviously useful for geodetic purposes such as determining the overall shape of the Earth and determining vertical crustal motion.

Subsurface density anomalies and changes such as those caused by water table, oil and gas reservoirs, mineral deposits,

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00057-9

man-made excavations, and even magma chambers underneath active volcanoes cause gravity changes of parts per million (milligal or mGal).

Smaller but still measureable gravity signals provide information about the motion of the Earth's crust due to convective processes inside the mantle of the Earth (Lambert et al., 1998).

Gravity measurement finds utility in scientific, commercial, and military applications. The scientific applications include many geophysical investigations such as determining the shape of the Earth (geodesy), Earth tides, plate tectonics (crustal motion), vulcanology, global sea-level changes, arctic ice-sheet mass balance, and water table variations, among many others. Many of the practical uses of gravity are described in Torge (2001). Gravity is also being used in the scientific field of metrology to help define a practical definition for the kilogram and to provide a link between the kilogram and the watt force balance (e.g., Newell et al., 1999, 1998). Commercial applications include oil/gas exploration, mining exploration, and reservoir monitoring. Military applications include navigation and detection of subsurface structures (tunnels or voids) (Edwards et al., 1997).

Many of these applications require a spatial measurement of gravity over the surface of the Earth, often referred to as a gravity survey. Gravity surveys conducted at one point in time over a physical area result in gravity snapshots or maps that provide information about deep subsurface density anomalies. Gravity collected over very large areas provides geodetic information about the shape of the Earth. Surveys over smaller areas can be used to infer subsurface density fluctuations and are useful for oil/gas and mineral exploration. Typical gravity surveys used for oil/gas exploration require a precision of about 0.1 mGal or 100 microgal (μGal).

Gravity surveys that are repeated over time are sometimes referred to as 'time-lapse gravity' or 4D gravity surveys (the latter term is borrowed from the seismic community where the four dimensions refer to the usual three spatial dimensions plus time). Time-lapse gravity surveys provide a much more sensitive way of looking at subsurface density changes due to fluid motion or man-made construction because large spatial gravity variations caused by deep structure that remains constant over time can be ignored since they are 'common mode.' Any static signature is removed when the time-lapsed gravity surveys are subtracted. Time-lapse gravity surveys are able to provide information at the level of a few parts in 10^9 of g even though the background static variations of gravity over the survey area are many thousands of times larger. The ability to measure slow (months or longer) nonperiodic gravity changes below 1 μGal becomes very difficult due to many environmental sources of gravity (air pressure, water table, etc.) changes, which are difficult to model. Thus, 1 μGal represents a practical limit for most terrestrial gravity measurements of slowly varying nonperiodic gravity signals.

The absolute gravity value at a given site is not usually needed to make a gravity map for commercial applications such as oil/gas exploration. In these cases, it is enough to simply measure the difference of gravity over a survey region. This is fortuitous because many gravity meters are able to detect small changes in gravity but have difficulty in determining an accurate value for absolute gravity due to excessive drift or tares between calibrations. Gravity meters that measure sensitive gravity differences in either space or time but are not able to give an accurate measurement of gravity are referred to as 'relative gravity meters.' Gravity differences over finite spatial distances and/or times are correspondingly sometimes referred to as 'relative gravity' changes. Most measurements of gravity employ relative gravity meters.

Relative gravity changes should not be confused with the subject of gravity gradients that are defined as the mathematical differential of gravity in all directions (a tensor with five independent components).

'Absolute gravity' refers to the exact value of gravity (connected directly to metrology standards of length and time) at any one point in space and time near the surface of the Earth. In most applications, the absolute gravity value is not needed but absolute measurements have the advantage that by definition they are calibrated and accurate.

Absolute gravity meters can be used in most applications, but often similar results can be obtained with a high-quality relative gravity meter. It is therefore difficult to make a hard rule about when absolute or relative gravity meters should be used but one can make a few generalizations. Absolute gravity measurements are often necessary for geoid (shape of the Earth) measurements or for determining absolute network points (used to calibrate relative gravity meters). Absolute gravity meters are also quite useful for measuring the slowly varying nonperiodic variations of gravity (over the periods of years). For example, absolute gravity has been used to watch small uplift signals (1 mm $year^{-1}$) in the Canadian shield due to postglacial rebound (e.g., Lambert et al., 2006) as well as for reservoir monitoring in the largest North American oil/gas reservoir in Prudhoe Bay, Alaska (e.g., Ferguson et al., 2008). Absolute gravity meters are required to calibrate relative gravity meters and for metrology work such as the redefinition of the kilogram in terms of the watt balance.

In most applications, smaller and less-expensive spring-based relative gravity meters achieve similar results in a more efficient manner than the more complex absolute gravity meter. Absolute gravity meters are being made smaller and less expensive, but most practical measurements of gravity will continue to be made with relative spring gravity meters. Spring gravity meters can be very small, light, easy, and quick to operate. They tend to be much less expensive than absolute gravity meters with equivalent precision.

There are special cases where high-precision relative measurements of gravity provide a lower-noise measurement for periodic signals than can be achieved with an absolute gravity meter. Monitoring normal modes of the Earth (Hafner and Widmer-Schnidrig, 2013) at periods of ~1 h after excitation by a large earthquake is a particularly important application requiring low-noise relative gravity measurements. The superconducting relative gravity meter is well known for its utility for these applications.

3.03.1.2 Definition of Precision and Accuracy

Before describing absolute and relative gravity meters, it is useful to discuss and define the common specifications of these gravity meters in terms of precision, repeatability, and accuracy. Loosely speaking, 'precision' is a measure of how fine a measurement can be made at any specific time and location.

Precision is often connected with the number of significant digits to which a value has been reliably measured. 'Repeatability' is a measure of the agreement between identical measurements but typically done at different times or after the measuring instrument has been moved and then brought back. The precision of a measurement can be better than its repeatability if the instrument setup or the passage of time affects the reading. 'Accuracy' is how many of the digits can be trusted to be 'absolutely' correct. The term accuracy typically relates to a reproducible standard usually kept by a national or international agency. These different terms are often used interchangeably by geoscientists (and manufacturers of instruments) when describing the quality of instruments or measurements, which can create confusion. It is therefore helpful to go a bit beyond these simplistic definitions and put these different measures of quality on firmer mathematical ground.

3.03.1.2.1 Precision

Precision is normally the least well-understood term for the quality of a gravity measurement, but it is one of the most important specifications for determining which instrument should be used for any given application. Precision refers to the relative uncertainty of a reported gravity measurement. It can be given in units of acceleration or as a fraction of the Earth's gravity value. For example, a 1 mGal precision can also be referred to as a 1 part per million (1 ppm) measurement of the Earth's field. The precision of gravity measurement can be thought of as the resolution or the smallest meaningful digit that can be determined from the instrument. Precision is a difficult quantity to understand because it depends upon four quite different things: the quality of the instrument, the noise of the site at which gravity is being measured, the length of time that the value is averaged, and the frequency band of the noise and/or signal. The precision can be limited by any one or by a combination of these factors.

Clearly, the quality of the instrument must be high enough to report a stable value over the time of the measurement. However, it can easily happen that the background noise (of a site) is higher than the fundamental or intrinsic noise of a gravity meter. In these cases, the background environmental noise will set the limit for the precision of the measurement. It is well known that high-precision gravity measurements are much easier to obtain in a deep underground mine far away from an ocean or man-made noise than on the busy streets of a populous city. The length of time for the measurement is also a critical parameter for the precision of any measurement. Averaging data is the most common way to increase the precision of a measurement. It is therefore always important to specify the averaging time associated with any precision specification. Finally, the noise often depends strongly upon the frequency at which the observation is made. For example, it is typically much easier to measure gravity signals that are varying over a few hours (e.g., tidal signals) than to measure very slowly varying signals (signals that vary over periods of months or years). This is because environmental gravity noise tends to have a red spectrum that increases at low frequencies. Often, it is impossible to separate the background and instrument noise, so it is important to remember that noise typically has several components.

The simplest type of noise is called 'white noise' and is characterized by a flat frequency (or amplitude) spectrum. Noise will have a white spectrum if all individual measurements are uncorrelated. In this case, then the precision at any given frequency will improve with the square root of the observation time. Another way of expressing this dependence on observation time is to define a spectral noise density, η, that, for gravity, has units of acceleration per square root of frequency (i.e., μGal Hz$^{-1/2}$). Since the width of each frequency bin is given by the reciprocal of the observation time, the noise in any one bin will shrink as the measurements are averaged. For a signal with constant amplitude and frequency, the signal to noise ratio will increase as the square root of the observation time as expected. The noise spectrum is a very good way to discuss noise even in cases where the noise spectrum is not white because although the noise density is a function of frequency, the amount of noise in each frequency bin still decreases with a longer observation time. The spectral noise incorporates the averaging time explicitly through its units. Figure 1 shows the measured gravity spectrum obtained from a spring gravity meter called the gPhone (serial number #106). The gravity data were sampled at 1 Hz and were collected for 1 day. It is clear that the noise between 0.05 and 1 Hz (1–20 s period) is dominated by microseismic noise and has a maximum spectral noise density of 100 μGal Hz$^{-1/2}$. This background noise is real and is caused by slow undulations of the ground at periods similar to typical ocean waves. It is interesting to note that these data were taken in Boulder, CO, very far from the ocean! At frequencies between 0.003 and 0.05 Hz (between 5 min and 20 s period), the noise level falls below about 2 μGal Hz$^{-1/2}$. It then climbs steadily higher at lower frequencies from 0.0001 to 0.001 (15 min–3 h); the gravity background rises rapidly due to the gravitational effect of the tidal forces from the Sun and Moon.

Much longer gravity records (years) have been made using superconducting (and other low-drift) relative gravity meters. The noise spectra from these long records show a very nice tidal band and then the noise rises at lower frequencies. Part of the rise at low frequencies is due to instrumental drift and some of it is due to very long-period Earth motion (plate tectonics, water table, etc.). A good discussion of the noise spectrum at long periods is given by Van Camp et al. (2005). Most of this

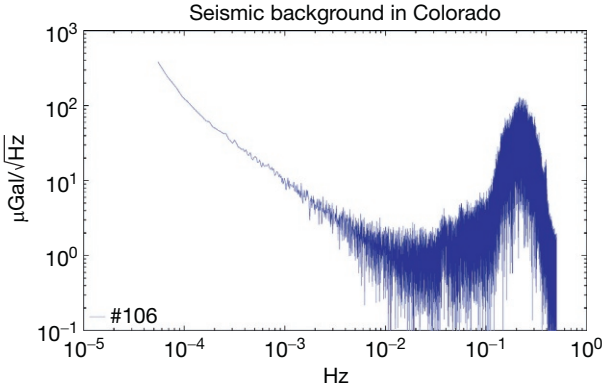


Figure 1 One day of 1 Hz measurements with a portable Earth tide gravity meter.

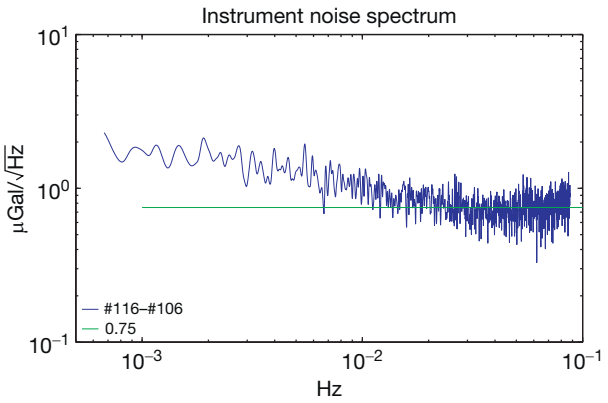


Figure 2 Spectrum of the difference between two gPhone gravity meters (#116–#106).

spectrum is dominated by real gravitational signals. At high frequencies, the signal is caused by seismic undulations of the Earth having an amplitude of about 1 μm at periods of 3–10 s. At lower frequencies, the spectrum is dominated by the tidal forces that oscillate with periods of 12 and 24 h as the Earth rotates causing variations in the local attraction of the Sun and the Moon on the gravity meter. The measured signal also contains any instrument noise.

One can untangle the background seismic noise from the instrument noise by subtracting two different instruments recording the same gravity signal. Figure 2 shows the instrument noise spectrum obtained by subtracting two identical gPhone gravity meters (#116 and #106) recording the same signal simultaneously at the same location. Real seismic noise and Earth tides are removed in the difference signal leaving the instrument noise behind.

The quietest frequency band between 0.001 and 0.09 is shown in the plot. In this band, the instrument noise level is about 0.75 μGal $\mathrm{Hz}^{-1/2}$. At higher frequencies, the noise increases due to timing errors between the two instruments that limit the level of cancelation of the large seismic noise amplitude. At lower frequencies, the noise increases due to instrumental drift of the instruments. Clearly for any measurements lasting less than an hour, the noise of the gPhone is dominated by seismic noise.

If the average background noise (seismic plus instrument noise) were white with an amplitude of η, the precision, P, of a gravity measurement would be given by multiplying the spectral noise amplitude by the square root of the smallest frequency bin. Alternatively, we can divide the spectral noise amplitude by the square root of the observation time given in seconds (i.e., $P = \eta/\sqrt{T_{\mathrm{observation}}}$), where η has units of $\mu\mathrm{Gal}/\sqrt{\mathrm{Hz}}$.

The noise background shown in Figure 1 is clearly not white and is dominated by noise in the seismic band 0.1–1 Hz. The data used for this spectrum were sampled at 1 Hz and had a standard deviation in the time domain of $\sigma_{1s} \approx 4$ μGal. The noise for the average value of the data will decrease as the square root of the length of the observation time. For example, 1600 s (about 30 min) would be necessary to achieve 1 μGal. This noise estimate and the corresponding measurement time are consistent with the experience of gravity surveyors who typically spend a few minutes with spring gravity meters at each field point. The experience with about 1000 surveys used to calibrate newly made CG5 gravity meters is that the repeatability is < 5 μGal.

It should be pointed out that the spectrum was obtained in a quiet location and a quiet section of the data was selected for the analysis. The background seismic noise can be much higher in cities or in active areas near the ocean. Higher background noise will increase the observation time needed for any given precision.

Any instrument, absolute or relative, must contend with background seismic noise. In general, the combined site and instrument noise is time-dependent. For example, it is quite usual in many different types of measurements that noise tends to increase at low frequencies (long time averages). Sometimes, this is called red or pink noise. A measurement that exhibits some sort of drift or linear variation of the measurement with time will give rise to a noise spectrum that increases at low frequencies. Drift in the instrument (or background) will cause the precision of a measurement to stop improving once the accumulation of the drift is equal to the uncorrelated noise component (which is diminished by longer averaging). Equating the random noise level with the drift allows one to calculate a critical observation time beyond which the decrease of random noise through averaging is offset by a larger drift error. Mathematically, this is given by an observation time, $T_{\mathrm{obs}} = [\eta/d]^{2/3}$, where η and d are the spectral noise amplitude and the drift of the gravity meter. For example, if we use a spring relative gravity meter with a drift of 1 mGal per month (30 μGal day^{-1}) and assume a high background (seismic) noise of 60 $\mu\mathrm{Gal}/\sqrt{\mathrm{Hz}}$, the crossover point will occur after about 1 h of observation. This is why it is useful to remove drift from the gravity measurements from spring gravity meters for surveys that last longer than about an hour. The drift of a spring gravity meter is normally estimated by repeating the gravity measurement at the same base point at the beginning and end of the survey in a process called 'looping.' This type of calculation can provide an unbiased way to choose the type of instrument needed and the optimal integration time for any given application.

Linear drift is the simplest type of time behavior for the noise of a gravity measurement. Most relative gravity meters exhibit an exponential drift that decreases over time and can be estimated as a linear drift for reasonably short observation times. Even very-low-drifting superconducting relative gravity meters are observed to exhibit an exponential drift when observed over many years (Van Camp and Francis, 2007). More troubling are instances where the value of gravity jumps (or tares) in relative gravity meters. Tares are extremely difficult to deal with because they cannot be predicted or removed with a model. Furthermore, one must recognize that the environmental background noise (real gravity changes that cannot be modeled) also contributes to the overall uncertainty and can even dominate the noise level of a measurement. At one extreme timescale (small timescales), one can consider gravity measurements taken on a boat. Over long time periods, the ocean surface (sea level) is quite stable, but at short time intervals, wave action can produce huge accelerations (100 000 mGal or more) that ultimately limit the precision of measurements. At long timescales (many years), the gravity

noise at a specific location on the Earth may become dominated by tectonic motion or water table variations. Large man-made structures (buildings and excavations) and activities (reservoir management) can also introduce measureable gravity variations. It is often difficult to distinguish between noise that is related to instrument imperfections and background environmental sources. Indeed, one can often obtain the same level of precision using different qualities of instrument when the environmental noise dominates the precision.

3.03.1.2.2 Repeatability

Repeatability is perhaps the most useful specification for a gravity meter because it can be used as a guide to what a typical user can expect to achieve in a real-world field measurement. Repeatability does not guarantee absolute accuracy but it does encompass other important sources of variation that may not be included in the reported precision of a gravity meter. All gravity meters exhibit some tendency to jump or 'tare' when they are moved. These tares can be a result of physical movement of the balancing mechanisms inside the instrument (similar to the tares found in a bathroom scale after moving). Tares can also be introduced by changes in the alignment of the gravity meter with the local vertical.

It is easy to understand why the common spring gravity meter is susceptible to tares when they are moved. Spring gravity meters are in principle very sensitive scales that balance the restoring force of a spring against gravity. The length of the spring and the mass must be controlled to an incredible level. For example, if the stretched length of the spring inside of a gravity meter is 1 cm, then the length of the spring must be controlled to 10^{-11} cm to achieve a precision of 1 μGal. This is ten times smaller than a molecule of water! It is easy to imagine that the length of the spring or the clamping mechanism of the spring may change by this small amount when the instrument is moved or bumped during a survey. Similarly, a small change of the balanced mass will also cause a tare in the measurement. The mass can change due to condensation caused by humidity or outgassing. Mass can also be removed or added by accidental rubbing of the test mass with the stationary parts of the gravity meter. Clamping the test mass to avoid damage during shipment is another activity that can change the mass being supported by the spring. Mass changes will cause the value of gravity to jump or tare.

In addition to tares internal to the gravity meter, there may be setup errors that introduce a small jump or tare every time the instrument is moved. For example, a gravity meter must be leveled to about 9 s of arc (~45 μrad) in order to achieve a 1 μGal reading. The error is quadratic in the absolute error in leveling. Leveling errors will normally cause a random jump in the measurements but will always be biased toward a smaller value of gravity because of the quadratic nature of the error.

The precision of the gravity meter together with drift, tares, and possible setup errors places a lower limit on the repeatability of any gravity meter. As always, environmental noise can also place a limit on the repeatability at any site if it is larger than the intrinsic instrument repeatability.

The repeatability for a survey gravity meter can be defined as the variation (usually 1σ) in gravity that will occur if the gravity meter is repeatedly set up at a given site, then packed up, and transported some distance and finally brought back to the same site. A common way to measure repeatability with a relative spring gravity meter is to make gravity measurements at several different stations and then repeat these measurements several times to check the repeatability of the gravity meter. These data can be least-squares fit to a solution that includes a unique value of gravity at every station and a drift parameter. The residuals of the fit can be viewed as a measure of the repeatability of the instrument. It is important to scale the standard deviation of the residuals by the square root of the degrees of freedom of the least-squares fit in order to arrive at the repeatability.

Figure 3 shows data from a CG5 spring gravity meter that was taken to a survey in Northern Canada in December 2003. The survey consisted of 39 stations on two different lines across the survey area. The average separation of each station was 40 m. The first line consisted of 23 sites (station ID #1–#23) and was repeated twice. The second line consisted of 16 stations (station ID #24–#39) and was also repeated twice. Station #1 was common for both lines and was repeated six times during the survey in order to control for drift. The survey locations are plotted in Figure 3 with the station ID given on the right-hand vertical axis. The gravity data were then least-squares fit to a

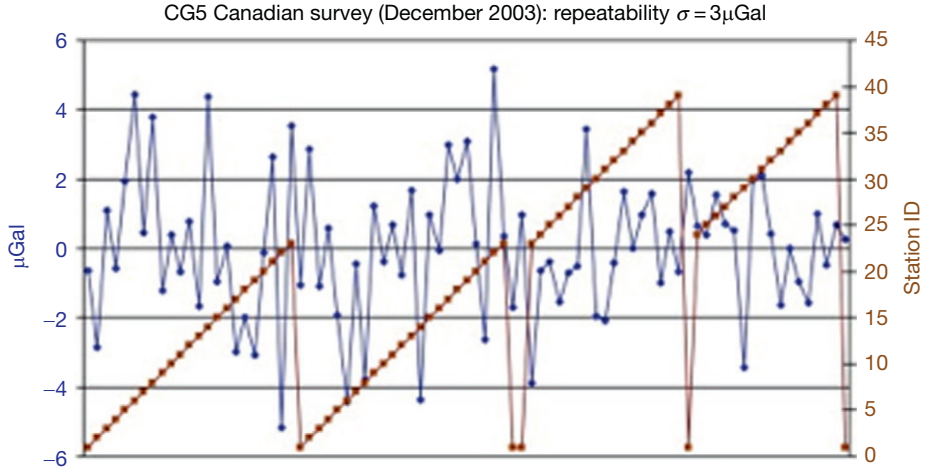


Figure 3 CG5 repeatability using data from a CG5 survey in Northern Canada during December.

network model consisting of 39 unique gravity values and a single drift parameter. The difference between the measured gravity and the model (residuals) is also plotted in Figure 3 using units given on the left-hand vertical axis. The residual difference has a standard deviation of 3 μGal. No tares or jumps in the gravity meter were observed during the survey. This survey is typical of data taken with the over 1000 different CG5 instruments that have been manufactured and is why the factory states that its repeatability is 5 μGal.

3.03.1.2.3 Accuracy

The accuracy of a measurement is much simpler to define and to understand than the precision or the repeatability. Unfortunately, accuracy is also quite elusive because there is no way to determine the absolute accuracy of an instrument unless a better instrument exists. Clearly, the accuracy of a gravity meter is limited by its repeatability and precision. In addition to the variability of the measurement, accuracy must also take into account any systematic errors (errors that do not change) that affect the reported value of gravity.

Accuracy refers to the correctness of the measurement in absolute terms. As we have discussed, gravity has units of acceleration given by meters per second per second (or $m\ s^{-2}$). These units are defined and kept by standards laboratories worldwide and are the concerns of metrology. The 'second' is defined as the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom at 0 K. The meter is now a derived quantity from the speed of light. It was redefined in 1983 to be the length traveled by light in vacuum during 1/299 792 458 of a second (Giacomo, 1984). In practice, the standard meter can be realized by an iodine-stabilized laser (Chartier et al., 1993) manufactured with a recipe given by the Bureau International des Poids et Mesures (BIPM) in Sevres, France.

An absolute gravity meter directly measures the acceleration of a free proof mass in terms of fundamental units of space and time as defined by the standards community. Spring gravity meters (including superconducting spring gravity meters) are not considered absolute gravity meters for two fundamental reasons. Firstly, they do not measure gravity directly but instead they measure the force required to keep a proof mass stationary. Secondly and more importantly, the measurement is not directly tied to metrological standards. The proof mass in a relative gravity meter is not tied to the kilogram and the force measurements usually are not tied to fundamental units such as the watt, meter, or second. Sometimes, the distinction between relative and absolute gravity meters is misunderstood to be associated with precision or repeatability. A common misconception is that absolute gravity meters are necessarily 'better' than relative gravity meters. In fact, it is quite possible to make an absolute gravity meter with much worse precision than a relative gravity meter. For example, early pendulum gravity meters, an early type of absolute gravity meter, had much lower precision than current relative spring-based gravity meters. At the other extreme, the relative superconducting gravity meter is the most precise instrument over some bandwidths but its accuracy is limited by the ability to calibrate the instrument *in situ* and long-term drift and/or tares.

Absolute gravity meters have one very demanding requirement. In order for an absolute gravity meter to be considered a metrological standard, there must be an associated systematic error budget that takes into account all errors in the instrument that can affect the measured value of gravity. It is notoriously difficult to identify, enumerate, and estimate the size for all errors in any instrument. Many systematic errors will not affect the precision or repeatability because they will remain fixed over time but they must be included in the systematic error budget. Some of these errors may be fixed with respect to a given instrument and some may depend upon the setup conditions or site conditions. The systematic error budget (Niebauer et al., 1995) for a modern-day free-fall absolute gravity meter called the FG5 has a limit of about 1–2 μGal.

It must be noted, however, that the accuracy of an absolute gravity meter cannot be independently checked except by other absolute gravity meters of equivalent quality or by a superior instrument. Different absolute gravity meters are sometimes brought to a single location to compare their measured values as check on the claimed accuracy of each gravity meter. International intercomparisons of gravity meters (Francis and Van Dam, 2003) show agreement between instruments of the same type to be consistent with the systematic error budget of the instruments. However, it is disconcerting that intercomparisons between different types of absolute gravity meters held at the BIPM in Sevres, France, every 4 years exhibit a variance in excess of 10 μGal (Vitushkin et al., 2002). Presumably, these differences indicate the presence of one or more systematic errors in one or more of the different instruments, but there is no simple way to determine which one has the larger systematic errors especially if the errors are constant and do not change over time. The disagreement (variation) between different absolute instruments accepted by the different standards laboratories in different countries sets how well a thing can be measured absolutely. The effort to refine absolute instruments and resolve discrepancies between different methods and absolute instruments for measuring gravity is an ongoing effort by the metrological community.

The situation where different types of absolute gravity meters are able to measure with a high precision and yet disagree with the values obtained with similar instruments is unfortunately a rather common occurrence in gravity. Gravity is the weakest force in the universe and is 10^{40} times smaller than electrostatic forces. This means that one must eliminate the many subtle influences of the nongravitational interaction of materials to an extraordinary level in order to measure gravity. A further difficulty with gravity is that it is fundamentally tied to acceleration through the laws of general relativity. According to Einstein, one cannot distinguish the difference between a gravity meter in an accelerating elevator and one that is stationary with a different gravity field. This means that any motion of the ground couples directly and fundamentally into the measurement. That is why gravity simply cannot be measured as well in an area with high seismic activity as it can at a very quiet site.

The difficulty associated with measuring gravity can be seen even more dramatically in the measurement of the fundamental gravity constant, G, often referred to as big G (the Earth's gravitational attraction is known as little g). G can be measured with a precision of 1 ppm or higher but apparently different

methods can disagree at the 1% level even when these measurements are made carefully by groups of scientists in the world's best metrology laboratories (Davis, 2010; Quinn, 2002). This situation can be frustrating for the scientist who needs to know G precisely, but it is a fruitful ground for the metrologist who wants to understand the source of systematic errors or for the scientist who is seeking for new information about the fundamental laws of physics.

Finally, it should be mentioned that many relative gravity meters also have an accuracy specification. By definition, they do not have fundamental or intrinsic accuracy because they do not measure fundamental quantities that can be tied to absolute standards. However, relative gravity meters gain accuracy after they are calibrated against an absolute. A relative gravity meter once calibrated can in turn calibrate other relative gravity meters as long as its calibration remains valid.

There are several ways to calibrate gravity meters. The most common method involves taking the gravity meter to a calibration range (available in most countries) with stations or sites having large altitude differences with correspondingly large gravity differences. Once these sites have been measured with an absolute gravity meter, a relative gravity meter can be calibrated by visiting each of them. The measured gravity differences can be compared to the absolute gravity differences to determine the proper scale factor to correctly convert spring stretch into a gravity difference (Valliant, 1991). It is quite easy to find sites in a mountainous area with gravity differences in excess of 100 mGal. Assuming that the gravity differences have been measured with a precision of 10 μGal by a relative gravity meter, a calibration factor can be generated with a precision of 10^4 (0.01% or 100 ppm).

Another way to calibrate a relative instrument is to measure the local tidal variations simultaneously with an absolute gravity meter and cross correlate the measurements to determine the calibration factor for the relative gravity meters. This method requires a much higher measurement precision (and thus a longer observation time) because of the comparatively small variation of the Earth tides that have amplitudes at 12 and 24 h of about 100 μGal, typically a thousand times smaller than what can be achieved by visiting sites with different altitude. According to Francis et al. (1998), a calibration precision of 10^{-3} (0.1% or 1000 ppm) can be achieved for a superconducting relative gravity meter using a 3–5 day simultaneous recording of the tides with an FG5 absolute gravity meter. Despite the lower precision, this method is important because it can be done without the need to move the relative gravity meter (Imanishi et al., 2002). This method is also useful even for gravity meters with high drift because all of the signal power of the tides is in a very narrow frequency band and drift is easily distinguished from tides.

It is possible to calibrate relative gravity meters without an absolute gravity meter by exploiting the equivalence of acceleration and gravity. A relative gravity meter can be placed on a moving platform and its motion can be measured in a calibrated way. The motion can be doubly differentiated to find the acceleration and then compared to the measured gravity difference.

The calibrations of dynamic gravity meters are routinely checked in this manner. A comparison of gravity measurements on a moving vehicle with the acceleration computed using differential global positioning system (GPS) data between a stationary antenna and one mounted near the gravity meter yields a calibration factor for the gravity meter. The tie to absolute quantities comes through the time and distance information broadcast by the GPS satellites.

Another way to calibrate a gravity meter is to place it on a platform designed to move vertically in a very precise and calibrated way (Richter et al., 1995). The platform must have a calibrated distance scale and the data must be taken with a calibrated clock.

Relative gravity meters are normally used only for measuring relative gravity differences between two points in time at the same location or two different locations. Thus, the calibration is done to establish a scale factor for measuring gravity changes and not typically for measuring the value of absolute gravity at a given point. Relative gravity meters do not have enough dynamic range and they have too much drift to make a single point measurement of gravity with nine or more significant digits. Absolute gravity meters are usually required for a single point measurement of gravity. Their dynamic range (nine or more digits) is possible because of the high fidelity of laser interferometry and accurate timing.

Despite the distinction being made between absolute and relative gravity measurements, it is important to emphasize that both types of measurements are useful. Absolute gravity is not needed for most real-world practical applications. As long as one knows how 'precise' and 'repeatable' their gravity measurements are, one can measure crustal motion and tidal signals as well as look for subsurface voids, or monitor earthquakes and reservoirs with a high degree of certainty with either type of instrument!

3.03.2 Gravity Meters

Gravity meters are separated into two categories: absolute and relative. Absolute gravity meters can be defined as any device or instrument that measures gravity directly in terms of standard time and length units ($m\ s^{-2}$). Early absolute gravity meters made use of the physical pendulum commonly described in elementary physics books in which the period is related to the location of the center of mass and gravity. In this case, the period and the distance can be linked to standards of length and time to give an absolute gravity determination. The pendulum was eventually replaced by a free-fall absolute gravity meter that monitors the descent of a freely falling body using laser standards and atomic time standards.

Relative gravity meters vary a great deal in the construction and the type of materials that are used, but basically, they are always a mass that is suspended by a spring. The spring material may be some sort of low-expansion metal or quartz. The spring can even be a superconducting magnetic field that produces a very-low-friction suspension. Relative gravity meters are distinguished from absolute instruments because they indirectly measure the acceleration of gravity by measuring the stretch of a spring. The stretch of the spring is proportional to a force constant (K = spring constant) that must be calibrated. The stretch is also proportional to the suspended mass. It would be theoretically possible to turn a relative gravity meter into an absolute gravity meter if the system could be

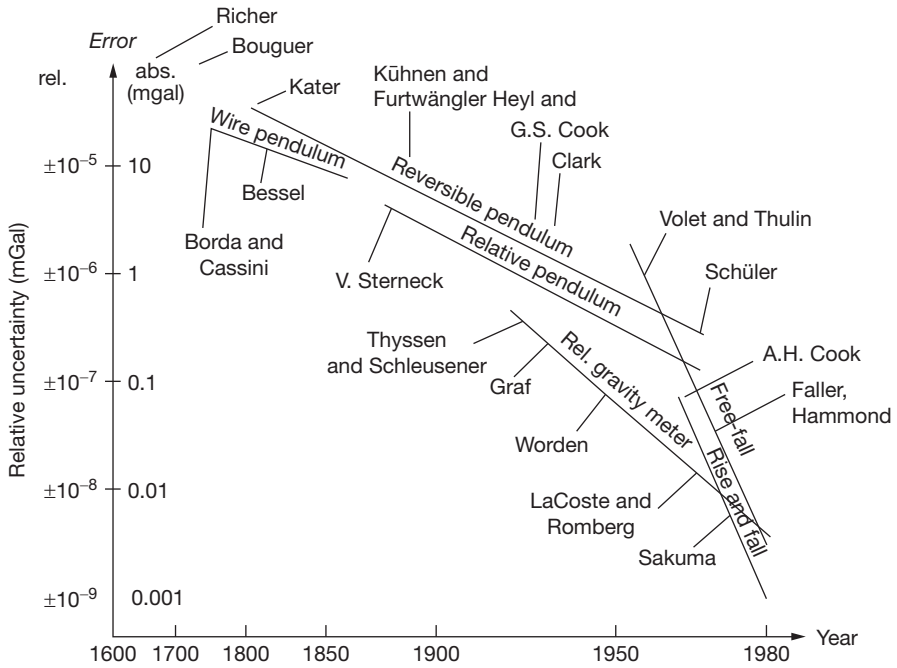


Figure 4 Accuracy of gravity meters, from Torge.

calibrated absolutely against standards. In that case, the test mass would have to be referenced to the standard kilogram and the force would have to be related to fundamental constants. This is not practical for existing relative gravity meters.

Figure 4 shows a historical progression of the accuracy/precision of gravity meters since the 1600 s. It is interesting to note that initially absolute gravity meters made with a pendulum were the most precise instrument until early in the 1900s where spring gravity meters jumped ahead of the pendulum. Free-fall gravity meters started behind the pendulum but quickly advanced to meet and finally beat spring gravity meters in the 1980s. The graph ends at 1980 largely because gravity measurements have a similar repeatability today (2013) of about 1 μGal. This limiting accuracy is mostly due to environmental sources of gravity errors and background noise level of the Earth. Higher-precision gravity measurements can be made but only for cases where the signals are periodic or have a known time signature and persist over long times so that the background noise can be averaged to obtain higher precision. For example, it is possible to measure periodic tidal signals with a precision of nanogal ($10^{-12}\,g$) if one observes gravity continuously for long periods of time (years). Long-period measurements (DC) and/or short-step measurements tend to be limited by the background noise level of 1 μGal (Van Camp et al., 2005).

3.03.2.1 Absolute Gravity Meters

There are two methods that have been used to realize an absolute gravity meter: a proof mass on a pendulum and a proof mass in free fall. Both methods allow a proof mass to fall under the Earth's gravity; the pendulum has a radial constraint while ballistic instruments use unconstrained free fall. The force of gravitational attraction of the Earth is proportional to the falling mass in both of these methods so that the resulting acceleration (given by the force divided by its mass) is independent of the size or shape of the falling proof mass as long as air resistance is minimized. One might suspect that the mass enters into a pendulum through its moment of inertia, but as we shall see, there is a clever method to avoid this problem. The mass used in absolute gravity meters is sometimes called a test mass or proof mass to highlight the fact that any mass would follow the same path with the same acceleration. Proof masses follow the same invisible geodesic in the same way that different leaves trace out the same path down a flowing stream regardless of their size. Free-fall absolute gravity meters can use macroscopic proof masses made of glass and metal, or they can even use individual atoms if they are properly cooled with lasers before they are allowed to free fall.

3.03.2.1.1 Pendulum absolute gravity meters

Galileo noticed that the period of a pendulum is nearly independent of the amplitude of its swing in 1602, which gave him the idea of using pendulums as clocks. Christian Huygens built the first successful clock based on the pendulum in 1657. The period is dependent upon the acceleration caused by the attraction of the Earth, '*g*,' so the pendulum can be used in reverse to measure gravity if an independent clock can be used to time the pendulum period. Gravity measurements made with pendulums were first made popular by Pierre Bouguer in 1749. The pendulum was, in fact, the primary method for measuring absolute gravity for about 200 years. A very accurate absolute pendulum gravity meter was developed by Captain Henry Kater (1818). Kater realized that a pendulum could be swung around two different axes in order to measure absolute gravity using a simple distance and time measurement. This method removed the dependence of the frequency on the suspended mass and replaced it by a single time and distance

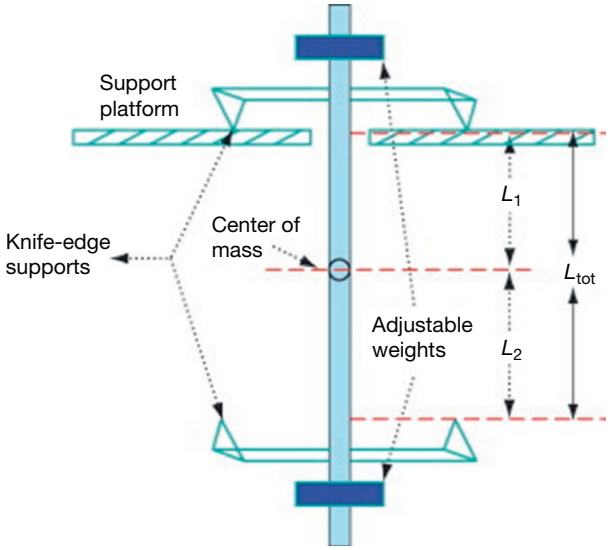


Figure 5 Schematic diagram of the Kater pendulum.

measurement that could be calibrated against the absolute standards of the day.

Figure 5 shows a schematic diagram of the Kater pendulum. It has two sets of knife edges that can be used to swing the pendulum from two different points about its center of gravity. The center of gravity can be adjusted using adjustable weights in order to make the swing period for the two suspension points equal.

The equation of motion for a pendulum (with no friction) can be written as

$$\ddot{\theta} = -\frac{mgL}{I}\sin(\theta)$$

where θ is the angle of swing and m is the mass of the pendulum. The distance between the knife edge and the center of mass is given by L, and the moment of inertia of the pendulum about the axis of rotation defined by the knife edges is given by I. In general, the distance between the center of mass and the knife edge and the moment of inertia will be different for the two sets of knife edges. The period is nearly independent of the swing amplitude as long as the angles are small. The period of the pendulum around each knife edge is given by

$$\Gamma_i = 2\pi\sqrt{\frac{I_i}{mgL_i}}, \quad i = 0 \text{ or } 1$$

The moment of inertia around each knife edge is related to the same moment of inertia around the center of mass by $I_i = I_{cm} + mL_i^2$. All moments of inertia can be eliminated by subtracting the two moments about the knife edges to obtain

$$m\left(L_2^2 - L_1^2\right) = \frac{mg}{4\pi^2}\left(L_2\Gamma_2^2 - L_1\Gamma_1^2\right)$$

Gravity is now given simply by

$$g = \frac{4\pi^2\left(L_2^2 - L_1^2\right)}{\left(L_2\Gamma_2^2 - L_2\Gamma_2^2\right)}$$

Notice that gravity is given only by distance and time measurements. It does not depend upon the mass of the pendulum! The expression is further simplified if one tunes the periods around each knife edge to be the same by moving the adjustable weights at the end of the pendulum. In this case, the difference of the two distance measurements is eliminated and only the sum of the two distance measurements becomes important. The expression for gravity then becomes

$$g = \frac{4\pi^2(L_1 + L_2)}{\Gamma^2} = \frac{4\pi^2 L_{tot}^2}{\Gamma^2}$$

This result is nice because one does not have to locate the center of mass because the only distance that matters is the total distance, L_{tot}, between the flexures. This equation shows that gravity depends solely upon two measurements one of which is a distance and the other a time. This is an ideal absolute gravity meter because one can obtain a calibrated result simply by measuring the distance between the knife edges with a calibrated standard meter stick and the period with a standard calibrated clock. The Kater pendulum is thus calibrated by its very design. This is a requirement for an absolute gravity meter.

The pendulum was a very successful absolute gravity instrument, but it finally reached its limiting accuracy of a few parts in 10^7 due to the difficulty in producing low-loss knife edges. The pendulum was eventually replaced with a free-fall gravity meter in which the proof mass was allowed to free fall in a vacuum. This eliminated the frictional problem inherent in the pendulum.

3.03.2.1.2 Free-fall absolute gravity meters

Free-fall or ballistic absolute gravity meters in which the proof mass is allowed to fall freely under the influence of gravity became technically feasible in the 1960s due to the availability of high-quality vacuum technology, electronics, and clocks. Obviously, there are many technical difficulties involved with dropping (or throwing upward) an object and then determining the distance and time for free fall. Currently, the trajectory of the falling mass is monitored using interferometry. Coherent light (typically from a laser) is split and part is reflected from the falling object, while the other beam reflects from a stationary object. The two beams are then recombined to form optical interference. The situation is similar to a demonstration that can be done using water waves that are split and sent in on two different paths before they are recombined. The waves can either recombine in constructive interference when the two waves are added trough to trough and crest to crest. Destructive interference occurs when the waves are added trough to crest and cancel each other out. As the proof mass falls a distance equal to half the wavelength of light, the optical interference repeatedly cycles through constructive and destructive interference (light and dark fringes). This provides a very finely calibrated 'ruler' with which to measure the height of the falling object.

There are a few important technical issues to solve before the acceleration of a freely falling proof mass can be measured. The proof mass must be launched vertically in order to keep it from rotating or translating so much that the reflected light can no longer be recombined with the stationary beam. One must also catch the object in a nondestructive way so that the measurement can easily be repeated. It is useful to automate the

system so that repeated measurements can be done quickly. A less obvious but important part of the absolute gravity meter is that it should provide isolation from ground noise or seismic surface noise that will otherwise impose a large background noise on each free-fall measurement, which lasts only about 0.1–0.2 s. The absolute gravity meter must have a very large dynamic range (10^9) in which all of the digits are useful and cannot be corrupted by systematic errors (due to air resistance or electromagnetic forces). Thus, it is very important to shield the proof mass from any force (other from gravity) that could inadvertently push on it during its free fall.

Many different types of absolute gravity meters can and have been built, but they generally consist of three independent systems:

1. Free-fall chamber
   a. A vacuum chamber to remove air resistance
   b. A proof mass that will be monitored during free fall
   c. Isolation from electromagnetic influences on proof mass
   d. A catch and release mechanism
2. Time and distance measurement system
   a. Laser interferometer that encodes free-fall distance onto the phase of an atomic system
   b. Wavelength-stabilized laser tied to metrological standards
      i. Mach–Zehnder or Michelson-type interferometer
      ii. Atom interferometer
   c. Precise time measurements of fringes made using an atomic clock standard
3. Seismic isolation for the stationary proof mass

3.03.2.1.2.1 The free-fall chamber

The free-fall chamber contains the catch and release mechanism for the freely falling proof mass. The trajectory of the proof mass should be sufficiently vertical to avoid Coriolis forces arising from the rotating Earth and also so that the light will continue to reflect off the object during its descent. The proof mass also should not rotate too fast so that rotational motion is confused with linear acceleration in the vertical direction.

The free-fall chamber should also shield the proof mass from electromagnetic forces. This task is a bit daunting given that on a molecular level, electrostatic forces are 10^{40} times larger than gravitational forces! The free-fall chamber is usually grounded and conductive so that it acts like a type of Faraday cage to keep outside sources of electrostatic charge from influencing the interior. The dropped object (proof mass) is usually fabricated to be conductive (to avoid static charges) and nonmagnetic. However, even a conductive nonmagnetic test mass will feel a magnetic eddy force when dropping through a magnetic field gradient. Care must therefore be taken to keep magnetic field gradients minimized as well.

The free-fall chamber is evacuated to avoid forces due to air resistance. One must eliminate these forces below 10^{-9} N if we want to measure the acceleration of a 100 g freely falling mass to 1 μGal precision. A nice experiment to demonstrate the dependence of air resistance on velocity is done with physics students at the Hockaday School. Figure 6 shows the measured force of air resistance as a function of velocity measured by dropping ten coffee filters (a 10 g mass) with a large cross-sectional area. Notice that the force is 0.1 N at about 2.8 m s^{-1} in air. This plot implies that one must decrease the pressure by at least nine orders of magnitude to get the air resistance low enough to measure absolute gravity with high precision. This means that a vacuum of about 10^{-7} torr (1 atm = 760 torr) would be needed for a 1 μGal free-fall measurement.

A similar limit has been determined (Niebauer, 1987) for a more conventional dropped object using a calculation of air drag (Kennard, 1938) given by $F_{drag} = \rho ScV/4$, where ρ and c are the air density and mean speed of the air molecules, S is the surface area of the dropped object, and V is its velocity.

The potential sources of systematic errors for measurements of absolute gravity have been described and cataloged (Niebauer et al., 1995) and are manageable at parts in 10^9 g and below if care is taken.

One particularly successful strategy to address all of these concerns was first perfected by Zumberge (1981). He used a small elevator to drop and catch a reflective proof mass. The proof mass terminated in three spherical 'feet' that rested in three V-grooves fastened to the floor of the elevator.

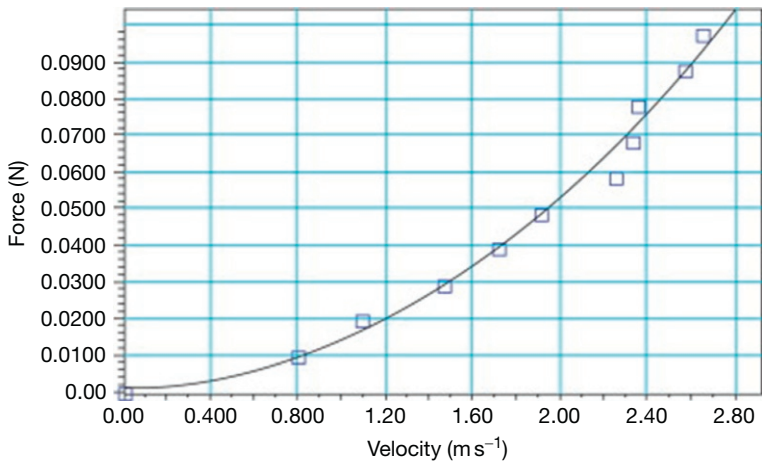


Figure 6 Air resistance versus velocity, from Hockaday.

The elevator was gently lifted to a predetermined height where it rested momentarily before the drop was initiated. Once the drop started, an external motor drove the elevator downward at about 2 g for about 20 m s until it separated from the proof mass by about 2 mm. Once the separation was achieved, the elevator reduced its acceleration to g in order to maintain a fixed separation between itself and the freely falling test mass during the measurement part of the free-fall trajectory. The proof mass and co-moving elevator fell about 20 cm in 0.2 s during which time the trajectory of the test mass was externally measured using a laser interferometer. Near the bottom of the free-fall chamber, the elevator acceleration was decreased until the object was caught softly (while still falling at close to g). Once the object was caught, the elevator was slowed quickly to a rest position at the bottom of the free-fall chamber. This cycle was automated and could be repeated every few seconds.

The co-moving elevator solved several problems at once. First, it provided an elegant way of dropping and catching the object in a nondestructive way that produced a release mechanism that did not impart very much horizontal or rotational velocity to the test mass. Second, it also acted as a sort of miniature Faraday cage to isolate the test mass electromagnetically from the rest of the free-fall chamber motor and vacuum pump voltages. In addition, it acted as a kind of 'drag-free' chamber to reduce air resistance on the test mass. The drag-free chamber reduces the vacuum requirement to about 10^{-6} torr.

Drag-free satellites (Lange, 1964) are a common method of avoiding degradation of low-altitude orbits. In a very real sense, the freely falling test mass and drag-free elevator become a very-low-orbit Earth satellite inside of an external drag-free satellite. The only difference is that the orbit of the proof mass inside the absolute gravity meter is tragically short-lived (0.2 s) and must be relaunched after each encounter with the Earth. We sometimes jokingly refer to this method as 'satellites in a can.' The advantage of this method over an actual satellite orbiting the Earth is that gravity is measured much closer to the Earth's surface and closer to any mass inhomogeneity inside the Earth. It is also clearly much cheaper to launch satellites in a can!

For completeness, it should be mentioned that the idea of dropping an object inside of a co-falling chamber was also used for earlier free-fall instruments built by Egorov and Martsinyak (1960) and Tate (1968).

Sakuma (1970) developed an important variant of the free-fall method in which the proof mass was launched upward, and measurements were made on both rising and falling portions of the trajectory. This method eliminates (or reduces) some classes of systematic error that tend to cancel on each half of the measurement. Air resistance, for example, is a downward force as the test mass travels upward but then becomes an upward force when the test mass is falling downward. On average, this force cancels and allows accurate measurements of gravity with a lower level of vacuum. Sakuma launched his proof masses upward by releasing an elastic element. This device was not automated and required a great deal of time to prepare so that only a few measurements could be made in a day. More recently, a co-moving elevator system was successfully used to launch a proof mass repeatedly in a rise and fall manner (at a rate of about two launches per second) at Micro-g LaCoste, Inc. This was not pursued, however, because the launch was found to introduce a horizontal velocity that grew in magnitude as physical wear accumulated on the contacts of the proof mass. This does not happen when the proof mass is dropped from rest because the initial velocity of the proof mass is so much smaller during release. While the rise and fall method continues to be a useful area for research, it has not yet become a practical method for any commercial absolute instrument.

It is also possible to drop cooled atoms instead of macroscopic objects. Kasevich and Chu (1991) at Stanford pioneered this free-fall method. They achieved an absolute measurement of gravity to a precision of 3×10^{-6} after a 1000 s integration time by monitoring laser-cooled sodium atoms in a free-fall chamber. Atom interferometry was used to determine the average free-fall acceleration of the atoms. They predicted that ultimately this technique would allow the absolute determination of gravity to 1 part in 10^{10}.

The atomic method has some advantages over dropping macroscopic mirrors whose position is monitored with laser interferometry. First, there are no 'moving parts' that can wear out. This has the potential of eliminating the need to service absolute gravity meters. Its other large advantage is that there is essentially no system recoil when the atoms are dropped. When a macroscopic proof mass initially at rest is dropped, there is a natural tendency for the system to rebound or recoil. Absolute gravity meters can employ a counterweight to reduce recoil (Niebauer et al., 2011a,b), but it is difficult to remove all impulses caused by the launch from the measurement system. Atom gravity meters do not have any issue with recoil. Many scientific groups are currently researching methods to improve atom gravity meters. Their current accuracy is very close to the ballistic gravity meter, but at the moment, they appear to be still about a factor of three to five times less accurate according to Louchet-Chauvet et al. (2011): "Yet, the differences between these different instruments and/or between different measurements realized by the cold atom gravity meter at the same station can amount to 15 μGal (1.5×10^{-8}), which is marginally compatible with the instruments uncertainties." Atom gravity meters have the potential to become the technology of choice in the future once the techniques have been optimized.

3.03.2.1.2.2 The distance and time measurement: Trajectory determination

The trajectory of a freely falling proof mass must be accurately determined in order to calculate the acceleration g (gravity). This is usually done with some sort of laser ranging. It is more challenging to do this for near-Earth satellites (satellites in a can) than traditional laser ranging for satellites orbiting high above the Earth because the orbit is very short-lived (0.1–0.2 s) inside the free-fall chamber. A very good method for measuring the trajectory of a test mass is to use an optical interferometer in which the free-fall distance is used to modulate the phase of light reflected from its surface. Faller and Dicke first pioneered this method in 1963 before the advent of common lasers using a white-light interferometer. **Figure 7** shows a schematic diagram of the apparatus. A monochromatic but incoherent (white) light source was collimated and fed into an optical

system (interferometer) that split the light so that some of it reflected off a falling retroreflector (sometimes known as a cat's eye) and another portion reflected from fixed mirrored surfaces. The light was then recombined onto a photodetector to convert the intensity variations caused by optical interference of these two beams into an electrical signal that could be used for timing. Incoherent light will only interfere when the two path lengths are identical. When this special condition is met by the falling retroreflector, the recombined light will momentarily be bright (or dark) due to constructive (destructive) interference. The interferometer was designed so that this special condition of equal path lengths between the light source and the falling test mass and the reference mirror occurred at three distinct displacements of the falling test mass. Figure 8 shows the three different cases or displacements of the falling test mass together with the corresponding fixed light path that caused a white-light interference. The time for each white-light fringe (or interference) was then measured. In this case, a fringe can be thought of as a momentary bright flash of light when the path length of the light reflecting from the falling object matches one of the three fixed light paths reflected from the stationary mirror. (It is common to use the word 'fringe' for the time-varying behavior of the recombined light intensity as the constructive or destructive interferences occur even though one usually thinks of a fringe being a spatial variation of intensity.) These three conditions provided three distinct time and distance pairs (t, z): $((0, 0), (t_1, h), (t_2, 2h))$ that were sufficient to solve for the initial velocity and acceleration of the object. The acceleration is independent of the initial velocity, which means that the precise starting position of the drop was not critical to the measurement of gravity. The value of gravity is also independent of the composition of the dropped object because all matter experiences the same force of gravity according to Einstein's equivalence principle.

The white-light interferometer was quickly replaced with a coherent laser source once it became available. A coherent laser produces fringes in an interferometer even when the two arms do not have equal paths. Constructive interference occurs whenever the path length of the reflected beam (the changing arm of the interferometer) differs from that of the reference path (fixed arm) by any multiple of $\lambda/2$, where λ is the wavelength of the laser. The factor of $1/2$ occurs because the path length of each arm is twice the separation between the beam splitter and the reflective component that terminates each arm. A laser removes the need to have three different fixed paths and allows the gravity meter to use a much simpler Michelson-type interferometer. A simplified Michelson-type interferometer is shown in Figure 9.

The mirrored surfaces are meant to indicate a retroreflector commonly called a corner cube (or cube corner depending upon which English variant is used). The corner cube has the important property that the reflected path length is independent of a horizontal translation or a rotation of the cube around the vertex. The corner cube also reflects the light parallel to the incoming beams so that the interferometer will stay aligned even if the falling corner-cube suffers small horizontal translation or rotations during its descent. These are very important practical advantages because it is not trivial to release or launch an object without imparting some initial horizontal velocity or rotation. In fact, the current free-fall method would probably not be viable without a corner cube!

The optical laser interferometer is an ideal method for measuring distance in a gravity meter because it has a very large dynamic range. The interferometer can be used to measure a free-fall distance of many meters or indeed many kilometers in the case of space satellites with a precision of under 1 nm. The corner cube permits the selection of only one axis of

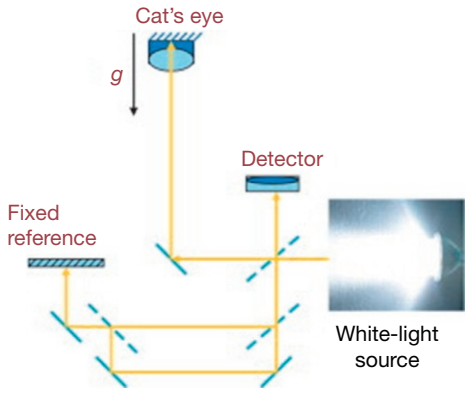


Figure 7 Schematic diagram of the Faller–Dicke absolute gravity meter (Princeton, 1963).

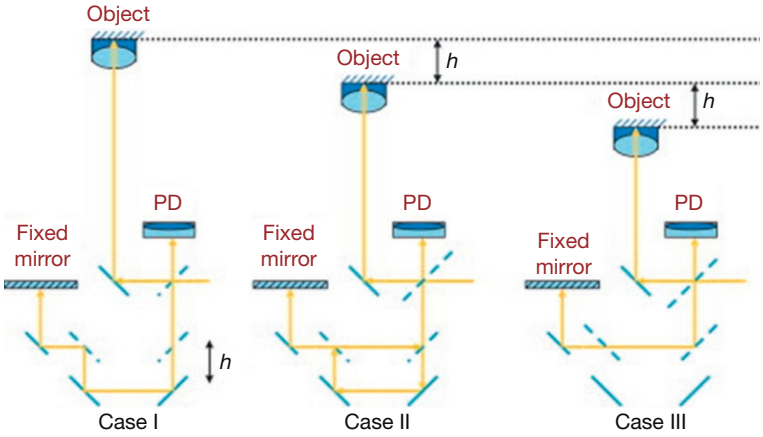


Figure 8 Three optical paths (cases I, II, and III) for interference.

motion with a high degree of rejection of motion in other axes (horizontal motion and rotation around the optical center). It is also possible to use other types of interferometers such as the Mach–Zehnder interferometer first used in the commercial FG5 absolute gravity meter. Figure 10 shows a schematic diagram of a Mach–Zehnder arrangement where the laser is split and one beam (the reflected beam) is sent into a changing arm that reflects from both a falling mirror and a fixed reference mirror. The other arm (transmitted beam) is combined with the changing arm on a second beam splitter.

It is possible to recombine both beams on the first beam splitter but then one must consider the multiple beam interferences that occur due to light traveling multiple times in the changing arm. In this case, the interferometer behaves more like a cavity. The variation with two splitters avoids cavity effects and results in simple two-beam interference. Two-beam interferometers can be understood simply as the recombination of two waves from each arm of the interferometer as shown in Figure 11.

When the two wave fronts are in phase, they recombine to produce a large amplitude, which is known as constructive interference. At this point, the light hitting the photodiode is bright. When the two waves are completely out of phase, the interference is destructive and the intensity of the light hitting the photodiode is dark. This pattern of light and dark fringes repeats as the object falls. The time for successive cycles for each fringe becomes shorter and shorter as the object falls faster and faster due to the acceleration of gravity. The optical signal that hits the photodiode produces a sinusoidal signal whose frequency increases linearly with time as shown in Figure 12.

The times for each zero-crossing of each fringe are denoted by t_n. This figure shows only a small fraction of the fringes that occur in a real gravity meter. In reality, the fringes start at DC and sweep to about 6 MHz over about 0.2 s during which time over 600 000 cycles or fringes are observed. The laser interferometer is similar to the white-light interferometer example given first except that now it is possible to make many more than three time measurements during the free fall. The frequency at any point in time corresponds to the instantaneous velocity, $f = 2gt/\lambda$, of the proof mass. The frequency increases linearly with time for a constant gravity value, g.

A sinusoidal signal with a linearly increasing frequency is sometimes referred to as a frequency chirped waveform. The chirp rate, given by the rate of change of the instantaneous frequency, is proportional to gravity.

A simple way to analyze this signal is to consider times when the sinusoid crosses its mean value. If the detected signal is 'AC-coupled' or high-pass-filtered, then the mean value is zero volts and these times are referred to as 'zero-crossings.' Zero-crossings occur at equal phase intervals in the interfering light that occur whenever the proof mass has fallen an integer number of $\lambda/2$ distance increments.

Figure 13 is a graph of the position and time for the first few zero-crossings.

One can choose some number of fringes (minimum of 3 and maximum of 600 000 for a 0.2 s drop using red laser light) and fit these to the usual curve for a freely falling body. If we assume that we measure the time for every M zero-crossings, the free-fall equation becomes

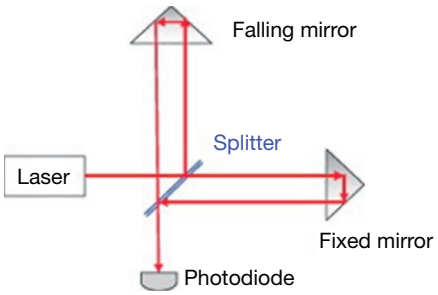


Figure 9 Michelson interferometer used to measure gravity from a falling mirror.

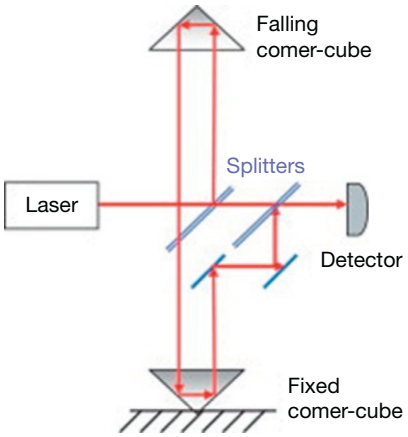


Figure 10 FG5 Mach–Zehnder interferometer for measuring gravity.

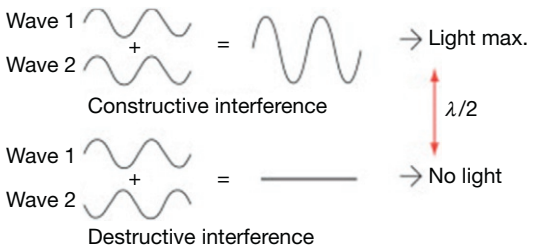


Figure 11 Two-beam interference with coherent light.

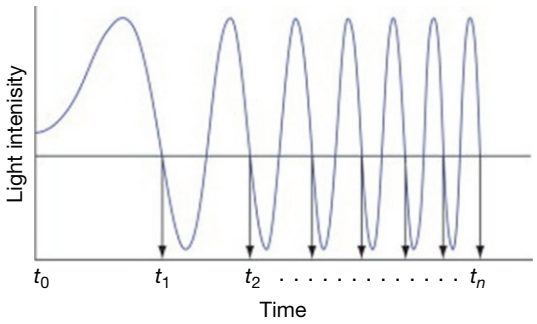


Figure 12 Two-beam interferometer signal with zero-crossings.

$$x_n = n\frac{M\lambda}{2} = x_0 + v_0 t_n + \frac{1}{2}gt_n^2$$

where the initial position, velocity, and gravity are linear parameters that can be determined using a common least-squares fit. This equation is sufficient to achieve a 0.1 mGal measurement for g. For higher accuracy, one must take into account the effect of the vertical gravity gradient, γ, in the equation that adds a cubic and quartic term as given by

$$x_n = \frac{M\lambda}{2} = x_0 + v_0\left(t_n + \frac{1}{6}\gamma t_n^3\right) + \frac{1}{2}g_0\left(t_n^2 + \frac{1}{12}\gamma t_n^4\right)$$

where x_0, v_0, and g_0 are the position, velocity, and gravity at $t=0$. This equation will achieve a 10 μGal measurement. In order to reach 1 μGal, one must take into account the finite speed of light. This is normally done by 'retarding' the time variable to take into account that the object position occurs at an earlier time than the fringe formation in the interferometer by an amount $t_n' = t_n - c/x_n$.

The method of zero-crossings is very convenient because it can be analyzed with a least-squares analysis linear in the gravity parameter. The other advantage is that one does not need to fit every zero-crossing. In fact, in the JILAg series of absolute gravity meters (Niebauer, 1987), only about 100 equally scaled zero-crossings for the data analysis. Even in current absolute instruments, only about 1000 zero-crossings are typically used.

It is possible to digitize the entire chirped waveform (not just the zero-crossings) and analyze these data using a nonlinear least-squares fit using parameters that can be determined in the time domain (Parker et al., 1995) and/or the frequency domain (Niebauer et al., 2006). This approach is becoming more and more popular as the speed of digitizing electronics and computation is increasing and memory has become much less expensive.

The value of gravity obtained with a ballistic free-fall type of instrument does not depend upon the initial position or velocity of the freely falling body and only depends upon the wavelength of light and the time measurements. The free-fall gravity meter produces a value for gravity that is 'absolute' or intrinsically calibrated by the laser wavelength and atomic clock used to make the measurements. For highest accuracy, a primary standard laser can be used for the light source and an atomic clock for the timing. The ballistic absolute gravity meter is a primary method for measurement of gravity for metrological standards.

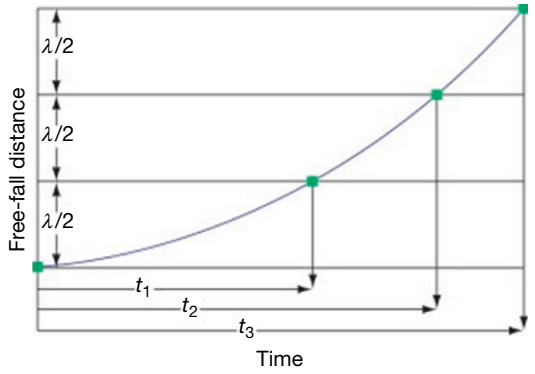


Figure 13 First four zero-crossings of the free-fall gravity meter (position vs. time).

3.03.2.1.2.3 Seismic isolation

Absolute gravity meters must measure the free-fall distance to about 1 nm precision in order to achieve a 1 μGal. Seismic noise varies around the Earth but generally is about 10^{-6} m $Hz^{-1/2}$ and is found to have a peak at a period of about 6 s (Raab and Coyne, 1997). Background noise therefore places a limit on precision for absolute gravity meters of about 100–1000 μGal $Hz^{-1/2}$ unless some type of seismic isolation is done. The FG5 absolute gravity meter accomplishes this by placing the reference arm of the interferometer on a long-period seismometer called a 'Super Spring' (Rinker, 1980). The Super Spring has a period of 60–90 s and can reduce the seismic noise so that the noise on one drop is about 3–5 μGal at a quiet location. It is also possible to measure the seismic noise using an accelerometer and then remove the ground noise from the measured free-fall signal.

3.03.2.1.3 Commercial absolute gravity meters

Currently, the most accurate commercial absolute gravity meter is Micro-g LaCoste's model FG5X. It is primarily a laboratory instrument consisting of an evacuated dropping chamber, interferometer, Super Spring, interferometer, and an iodine-stabilized laser (usually a WEO model 100 I_2 stabilized laser). The instrument can be assembled in about 1 h and is ready to take measurements once all of the components have reached thermal equilibrium. A photograph of the FG5X is shown in Figure 14. The instrument is about 1.5 m tall and the instrument requires about 1 m^2 of reasonably flat area for setup. At a quiet site, it is capable of obtaining a 3 μGal measurement after only one drop, but more typically, it has a single drop scatter of 10 μGal. Different FG5 instruments are consistent with each other at the 1–2 μGal level. This value is

Figure 14 FG5X absolute gravity meter, from Micro-g LaCoste, Inc.

consistent with the systematic error budget (Niebauer et al., 1995) and has been verified in several different intercomparisons (Francis and van Dam, 2006; Sasagawa et al., 1995) where metrology experts check instrument alignment and maintenance and then carefully set up the instruments. The positions of the instruments are interchanged to remove uncertainties due to horizontal gradients. The best intercomparisons are done at seismically quiet locations. It is interesting to note that when different types of absolute gravity meters are intercompared at a more active site and less control of the setup and maintenance of the instruments is required, errors of 10 μGal or larger are observed (Vitushkin et al., 2002). In general, exquisite care is necessary to achieve accuracies of 1 μGal with any type of gravity meter.

3.03.2.2 Relative Gravity Meters

There have been many different types of relative gravity meters developed over the years. Herschel proposed the use of a static spring to balance a mass against the force of gravity in 1833. Static spring gravity meters began to replace pendulum gravity meters from about 1930 onward. McGarva Bruckshaw (1941) provided a very detailed discussion of the many ideas that had been tried by 1941. Bruckshaw pointed out that the main advantage of the static gravity meters was that they could provide a measurement in a few minutes instead of an hour or two that was more typical for pendulum measurements. It is interesting that although Bruckshaw discussed many different gravity meters of his day during his lecture at the Imperial College London, he failed to mention the one gravity meter invented by LaCoste and Romberg in 1932 that had already made them all obsolete!

It is useful to point out the immense difficulty of measuring gravity with a spring before discussing some of the solutions that have been wrought over time. The spring extension for a simple vertically hanging spring is proportional to gravity. Thus, one must measure the extension of a spring with the desired precision of the gravity measurement. For example, if we consider a simple vertical spring stretched by 0.01 m (1 cm), its extension must be measured with a precision of 1 Å (or 10^{-10} m) in order to measure gravity to 10 μGal. One angstrom is the size of a very small atom (hydrogen). A human hair on this scale is 1 500 000 Å! It is obviously very difficult to measure the extension of a spring to this level even with today's technology. Even if the spring extension can be measured to the required precision, its length must be stable as time passes and with changes in temperature, humidity, or pressure. For example, the temperature dependence of the spring constant of most metal springs is about 10^{-3} per °C, which means that the temperature of the spring/mass system must be held to a constant temperature to about 10 μC if one wants to make a 10 μGal measurement with a common metal spring. The spring gravity meter is a great example of an instrument that works much better in practice than could be expected on simple theoretical grounds!

In 1932, Lucien LaCoste was given a homework problem to design a vertical seismograph by his professor Dr. Arnold Romberg. LaCoste's solution became a novel beam suspension that eventually became the cornerstone of a very successful line of spring gravity meters. His suspension amplified the stretch of the spring associated with gravity changes so that gravity could be measured optically with high precision before the advent of sensitive electronics. LaCoste and Romberg later founded the famous LaCoste & Romberg (L&R) gravity company in 1939 that revolutionized relative spring gravity instruments (Chris et al., 1996).

3.03.2.2.1 The L&R zero-length spring suspension

The L&R suspension is shown in Figure 15. A mass is suspended horizontally from a beam that can rotate freely around a low-friction pivot point. The force of gravity is balanced with spring that is somewhat disadvantaged by pulling upward at a nominal angle of about 45° (instead of pulling directly upward on the mass). As we will show, this suspension produces an astatic type of gravity meter (Heiskanen and Vening Meinesz, 1958), which, when balanced, results in no torque on the beam at any angle. The system becomes highly unstable when gravity is not perfectly balanced by the spring. This astatic system provides a great deal of amplification that was very useful for measuring gravity especially before sophisticated electronics became available. The system is therefore very sensitive with respect to gravity changes.

The astatic nature of the suspension can be seen by considering the potential energy of the system given by

$$U = \frac{1}{2}k(S - S_0)^2 - mga\sin(\theta)$$

where S is the stretched length of the spring and S_0 is the unstretched length of the spring (i.e., the stretch at which the restoring force from the spring is zero). The stretched spring length can be written as

$$S^2 = y^2 + b^2 + 2yb\sin(\theta)$$

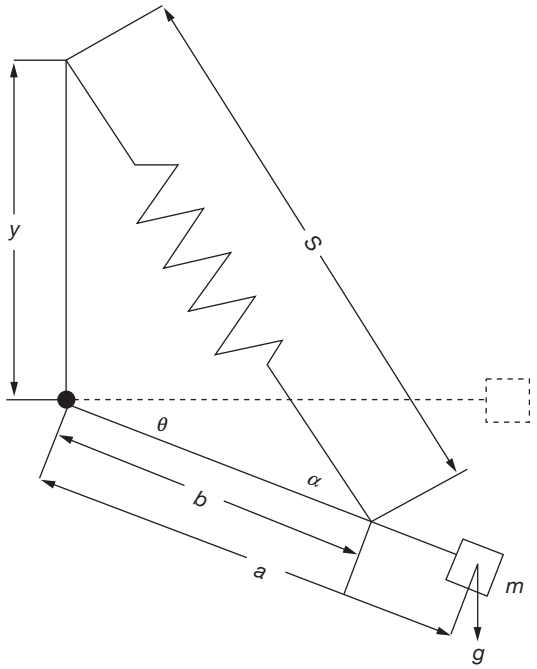


Figure 15 LaCoste and Romberg spring gravity meter suspension.

One can see immediately that if the unstretched length is zero (i.e., $S_0 = 0$), then the potential energy can be written as

$$U = 1/2k(\gamma^2 + b^2 + 2\gamma b \sin(\theta)) - mga \sin(\theta)$$
$$= 1/2k(\gamma^2 + b^2) + (k\gamma b - mga) \sin(\theta)$$

The gravity meter is balanced when $k\gamma b = mga$. Note that the balanced beam has a potential energy with no angular dependence. It follows that there will be no torque on the system at any angle, and therefore, there will be no preferred angle of the beam. Typically, the gravity meter is balanced by moving the top suspension of the spring up or down (thereby changing γ). Gravity is then directly proportional to the position of the top of the spring, γ, and can be written as $g = k\gamma b/ma$.

The torque on the beam, without any simplifying assumptions, is generally given by

$$\Gamma = -\frac{\partial U}{\partial \theta} = \left(mga - k\gamma b + \frac{kS_0\gamma b}{\sqrt{\gamma^2 + b^2 + 2\gamma b \sin(\theta)}} \right) \cos\theta$$

Again, we see the simplification if the unstretched length of the spring is zero ($S_0 = 0$). This removes the last term that depends upon the angle of the beam in a complicated way and gives a very simple form for the torque on the beam $\Gamma = (mga - k\gamma b)\cos\theta$. One can imagine how happy the eager student, LaCoste, was to set this term to zero in his homework problem and show the simplified torque formula!

The name, *zero-length spring*, was given to the special condition where the length of an unstressed spring is zero. A simple example of a zero-length spring is a flat circular unstressed spring. When the weight is applied to the center, the spring deforms creating a force that is proportional to the stretch of the spring from its zero position. An extension spring can also be made to behave mathematically as a zero-length spring as long as its pretension is equal to the force required to stretch the spring to its length with no load attached. For example, if a spring has a force constant $k = 1\ \mathrm{N\ m^{-1}}$ and an unstretched length of 10 cm, it would need to be wound with a pretension of 10 g in order to have zero-length. The coils of this zero-length spring will just start to separate after a weight of 10 g is placed on the end of the spring. There is some confusion in the literature surrounded by the advantages of using a zero-length spring. It is sometimes said to be the source of a long period, but in fact, its primary purpose in the L&R suspension is to create a beam suspension in which the restoring force is not dependent upon the beam angle.

Another common misunderstanding of the L&R suspension written in many books and papers is that the zero-length provides an infinite period spring suspension resulting in high sensitivity for gravity. While it is true that a spring with an infinite period also has no torque at any angle, this is not quite the same thing as an astatic beam. The difficulty with interpreting the L&R astatic beam as having an infinite period is that one typically associates a spring period with a stable system with some restoring force that becomes smaller and smaller as the instrument is tuned to have an infinite period. The L&R suspension, however, has a torque proportional to the cosine of the angle when the beam balance is not perfect (or $mga \neq k\gamma b$). This means that the beam will either fall down or up depending upon the relative strength of the spring and gravity forces when the beam is not perfectly balanced. In other words, the perfect L&R suspension is actually unstable as shown in Figure 15. The balancing condition is a saddle point rather than a point of stability. It is therefore mathematically incorrect to ascribe a period (even infinity) to this equation. The cosine dependence of the torque upon angle makes the system much more sensitive to small gravity changes than a spring with a truly infinite period. The likely criticism of LaCoste's professor Romberg to seeing this equation for the first time might have been to ask, "How can one use a system that is inherently unstable in a practical device?" The answer requires a reanalysis of the system where the attachment point of the spring is not directly above the pivoting hinge. This can be accomplished by tilting the instrument by an angle ϕ as shown in Figure 16.

Recomputing the torque for a zero-length spring ($S_0 = 0$), we find

$$\Gamma = (mga - \gamma kb\cos\varphi)\cos\theta - \gamma kb \sin\varphi \sin\theta$$

The last term is proportional to the negative of the sine of the beam angle. This term provides the necessary restoring torque to stabilize the beam against mismatches in the balance of the gravity meter. This tilt is necessary to bring stability to the L&R suspension. However, only a very small tilt ($\varphi > \theta$) is needed to stabilize the system once the beam is balanced. To first order in phi and theta, the balancing of the meter is still performed by adjusting $\gamma = mga/kb$ and the resulting torque is then given by $\Gamma = -\gamma kb\varphi\theta$.

The dynamic equation for the beam angle is then given by the classic harmonic oscillator:

$$\ddot{\theta} = -\omega^2\theta, \omega^2 = \frac{\upsilon kb\phi}{I} = \frac{mga\phi}{I}$$

If we have a massless beam, the frequency is given by

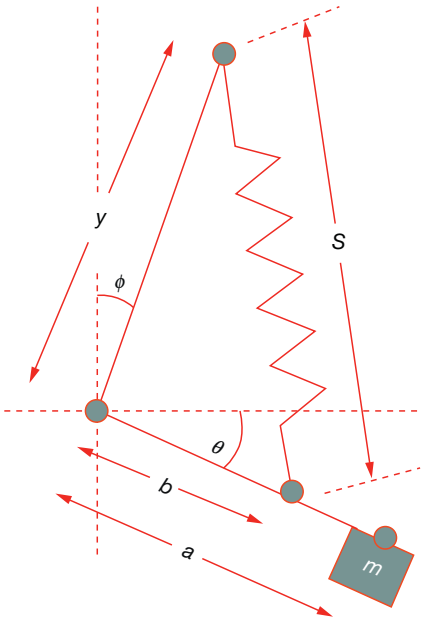


Figure 16 Tilted L&R suspension.

$$\omega^2 = \frac{mga\phi}{I} = \frac{g\phi}{a}$$

A tilt of $\varphi \approx 10^{-4}$ will provide a period of about 20 s for a 1 cm beam ($a = 1$ cm). This is equivalent to about 20 arcsec. Once the beam is balanced, a small change in gravity will tend to unbalance the beam, and it will sag until the restoring torque from the sin(θ) term balances the cosine(θ) term. This can be simplified to yield the equation $\varphi\theta = \Delta g/g$. Note that if Δg is 10 μGal and $\varphi \approx 10^{-4}$, then the beam will sag 20 arcsec (or $\theta \approx 10^{-4}$). Twenty seconds of arc was easily visible in the standard field telescopes available in the 1940s. This example shows how the L&R suspension allowed the use of a simple spring to measure gravity with high sensitivity even as early as the 1940s (before high-precision distance measurements and electronics were easily available). The dependence of the spring constant on temperature was overcome by using special low temperature coefficient spring alloys combined with a precise temperature control.

Figure 17 is a photograph of an L&R G meter. The instrument is leveled with the three black knobs. The top of the spring is moved with the silver rotary knob on the center of the top lid. The eyepiece of the telescope is also visible on the top of the lid. The telescope is used to look at the beam position. The instrument weighs 10 kg (meter, battery, and case) and has dimensions of $19.7 \times 17.8 \times 25.1$ cm. This gravity meter has been a popular survey instrument for over 50 years because it is lightweight and can take a 10 μGal measurement in a few minutes. Surveyors can collect as many as 100 points a day with this instrument in difficult field terrain. This instrument dominated land gravity measurements from 1940 and is still used today.

Figure 17 L&R G meter.

Electronic feedback was an important improvement in the L&R style gravity meter made possible with the advent of modern electronics. A quartz spring gravity meter was made using the L&R astatic suspension at Princeton in 1965 (Weiss and Block, 1965). This meter used an electrostatic force feedback to keep the beam at a null position. A voltage was applied to capacitor plates one of which is on the beam of the gravity meter to add a force feedback on the beam. This method removed the need to use the screw mechanism to increase the spring tension as long as the gravity changes were within the feedback range of the electronics. This was the beginning of a new era in spring gravity meters where the gravity meter beam is held fixed by electrostatics and gravity is read out as the force keeping the beam nulled. This technique holds the spring and beam geometry fixed even when gravity changes, which greatly reduces the linearity demands on the sensor over its measuring range. These feedback circuits were continually improved and were installed on L&R metal spring gravity meters (Block and Moore, 1966; Hugill and Valliant, 1986; Moore and Farrell, 1970).

The problem with the L&R meters was that the spring and test mass were too heavy to use feedback over the whole worldwide range of gravity values. Thus, it was still necessary to have a screw to adjust the spring tension whenever the gravity meter was out of range of the electrostatic feedback. For applications like static monitoring of Earth tides, it was possible to use only the electrostatic feedback, but for applications where the gravity meter had to move to different locations, the meter was used in a hybrid mode partially using the screw to range the gravity meter and the electrostatics to make a fine adjustment.

3.03.2.2.2 Quartz vertical spring gravity meters

The CG3 quartz gravity meter developed by Hugill (1990) at Scintrex Ltd. in 1990 had a very light test mass and was able to use electrostatic feedback to null the gravity meter over the entire worldwide range. Seigel et al. (1993) provided a more complete description of this gravity meter. The CG3 used a vertical quartz spring element rather than the L&R beam suspension where the spring has a 45° angle with the beam. A simple vertically hanging spring gravity meter requires a much better measurement of the vertical position (10 000 times better) than the L&R suspension because it is not astatic. The capacitive feedback system was similar to that used for L&R meters. Capacitive plates can achieve very high sensitivity to position having a noise floor of about 1 nm $Hz^{-1/2}$ at room temperature. The sensing usually is done with a three-plate sensor where a voltage is placed on the outer plates and the center plate is mounted at the suspended free end of the gravity meter. The voltage difference between each external plate and the center plate is measured. Typically, this voltage is modulated and synchronously detected for noise immunity. A DC force is applied to keep the position of the mass fixed when gravity changes. The quartz spring acts to levitate the mass, while the capacitive transducer is used to offset any changes in the gravity force on the mass. The same capacitor plates are used both as a position sensor and as a force transducer. The system is very insensitive to linearity of the spring constant because the spring is always kept at the same length by the capacitive electronics.

There are some advantages of using fused quartz over metal. Fused quartz is a very strong material, which has similar tensile strength per unit mass as music wire (piano wire) used for high-quality metal springs. However, it has a much lower coefficient of thermal expansion of about 5.5×10^{-7} K^{-1} (average from 20 to 320 °C), which is about an order of magnitude lower than for music wire.

Quartz is relatively easy to work using heat. This results in a monolithic sensor that avoids clamping of critical spring elements. The spring, the frame, and all connecting elements are made from quartz.

The current state of the art for quartz spring gravity meters is embodied in the CG5 commercially sold by Scintrex, Ltd. Its sensor is quite light and does not need to be clamped during movement. It has a monolithic quartz design that is much more resistant to tares that affect metal spring gravity meters of the same accuracy/precision.

3.03.2.2.3 The superconducting gravity meter

The gravity meter that used a superconducting magnetic field to levitate a mass was first developed by Prothero (1967) and Prothero and Goodkind (1968). The superconducting gravity meter consists of three parts: the levitated mass (sphere), the field coils, and the magnetic shield. The suspended mass is a 1 in. diameter hollow niobium shell manufactured with a slight mass asymmetry so that it has a preferred orientation when levitated. Two coils are used to independently adjust both the levitation force and the magnetic gradient. When current is placed in the coils, currents in the sphere are produced, which by the Faraday induction law precisely cancel magnetic flux from entering the sphere. The currents in the coils induce currents in the surface of the superconducting sphere that are perfectly stable in the absence of any ohmic losses. The interaction of the magnetic field in the coils and the sphere produces a 'springlike' force that is used to balance the gravitational force. The position of the mass (niobium sphere) is measured by a capacitance bridge that surrounds the sphere and is sealed with a partial pressure of helium gas in a separate cavity inside the coils. A superconducting cylinder with a hemispherical closure on one end provides the primary magnetic shielding. Currents induced in the shield's outer surface cancel any changes in external magnetic field due to the Earth or other nearby magnetic material. Current in the coils induces persistent currents in the inside surface of the shield, which prevent levitation magnetic field from extending outside the shield. An additional μ-metal shield is placed on the outside of the vacuum can. The entire instrument is cooled by liquid helium to 4 K.

The superconducting spring is of much higher quality than the metal or quartz springs discussed earlier. In principle, there is no temperature coefficient because it is held in a 4 K liquid helium bath. The principle of superconductivity produces a very uniform magnetic field free of gradients and other imperfections that would result from conventional room temperature coils. The drift from this type of gravity meter varies from 1 to 10 μGal year^{-1}. van Camp et al. (2005) recently determined that the superconducting gravity meter has a precision of 0.1–0.3 μGal Hz$^{-1/2}$. Thus, in most practical situations, the short-term precision is limited by the environmental noise background created by seismic noise.

The major difficulty with the current superconducting gravity meters is that they suffer a large uncontrollable tare when they are moved. This limitation may be removed in the future but at the moment this problem requires them to be used in static applications.

3.03.2.3 Gravity Meters Applications

The most common relative gravity meters are made using either metal or quartz springs and are much smaller and less expensive than either the absolute or superconducting gravity meter. These gravity meters are used for the bulk of all gravity measurements. Both types of spring gravity meters have similar accuracy and precision, but quartz gravity meters are resistant to tares at the 10–20 μGal level, whereas metal spring gravity meters typically exhibit tares of this size during a normal survey. Tares can be identified and corrected if the user is willing to make many loops to get repeated data at every site. Quartz gravity meters are therefore preferred on high-accuracy land surveys due to the less frequent tares at the 10–20 μGal. Some of the older metal spring gravity meters have the practical advantage in their ability to measure gravity without electronics due to the mechanical advantage provided by the lever arm and zero-length spring configuration. This can be helpful for surveys that are in very remote areas where an electronics failure cannot be tolerated. Either type of relative gravity meter can achieve a similar result if the measurements are taken carefully and enough repeat measurements are taken to identify and correct for drift and tares in the instrument.

Superconducting gravity meters are typically employed to measure Earth tides at a specific point on the Earth. It is common for them to provide amplitudes at the tidal frequencies with an incredible precision of 1 nanogal (0.001 μGal or $10^{-12}g$) after integrating the tidal signal for 1–3 years. Park et al. (2005) also showed that superconducting gravity meters were useful for observing free Earth oscillations excited by the Sumatra–Andaman earthquake in December 2004.

Some applications of relative gravity meters do not require them to be moved. They can be used a single site or at an array of fixed sites to monitor vertical crustal motions over the period of seismic periods (0.1–0.3 Hz), which can be used when studying signals originating from earthquakes (Niebauer et al., 2011a,b). There are also very interesting signals at longer periods (about 1 h) that correspond to the normal modes of the Earth's core (Hafner and Widmer-Schnidrig, 2013). These signals can provide information about the processes deep within the Earth that are responsible for plate tectonics. Tidal signals at 12 and 24 h are particularly strong and highly periodic signals that are easily detected and studied with gravity meters. Boy et al. (2003) used gravity records to compare different ocean loading models to provide better ocean models. At longer periods, gravity can be used to see water charging and discharging in reservoirs or storage facilities. Kennedy et al. (2012) used both relative and absolute gravity meters to monitor 8-week flooding cycles at each of three water basins.

Absolute gravity meters have traditionally been used for establishing base stations used for relative gravity networks. They are also used to calibrate relative gravity meters. Francis et al. (1998) described a method where a superconducting

gravity meter can be calibrated with an absolute gravity meter by correlating Earth-tide observations with both instruments over a few days of observations. Absolute gravity meters are particularly useful for monitoring long-term gravity changes (over 1 year or longer). The lack of drift and the good agreement between different absolute gravity meters allow measurements to be combined in a meaningful way to measure very small changes of gravity in real field conditions. Lambert et al. (2001) have taken absolute gravity measurements in Churchill, Canada, for more than 20 years (1987 to present) and have demonstrated that the Earth's crust is rising locally at about 3 mm year^{-1} due to postglacial rebound. These measurements provide new bounds on the mantle viscosity and provide an independent way to measure vertical crustal motion (other than GPS or VLBI).

The study of gravity changes is also known as 4D gravity in the commercial survey market. Absolute gravity measurements were started in 2002 and are currently being repeated annually in order to monitor subsurface density changes due to water injection into a large oil reservoir in Prudhoe Bay, Alaska (Brady et al., 2008). Approximately 300 stations are located over 10 km^2 area, many of which are on top of water in Prudhoe Bay. The survey is conducted while the surface of the bay is frozen so that high-accuracy static gravity measurements can be made. The survey required a special absolute gravity meter to be fabricated that could maintain temperature stability and be automatically aligned in temperatures that could be as low as −40 °C. **Figure 18** shows an absolute gravity meter deployed in Prudhoe Bay. The unit collects both gravity and location (latitude, longitude, and elevation) using GPS at every location.

The absolute gravity measurements are being used to monitor both the amount of subsurface water that is injected and the location of the flow. These gravity measurements improved the knowledge of the reservoir significantly over what could be measured by the sparse well information.

Borehole gravity also provides very useful measurements for resource exploration and for reservoir management. One obvious advantage of measuring gravity down a borehole is that the signals increase with the square of the distance to the subterranean resources. Gravity measurements down a vertical borehole have a surprisingly simple interpretation of lateral density variation for situations where the Earth density is layered horizontally. This follows from Gauss' law for gravity, $\vec{\nabla} \cdot \vec{g} = -4\pi G\rho$. This equation can be integrated over the volume Δof a right cylinder (a Gaussian pillbox) located at some depth in the Earth. With the assumption of lateral symmetry, the result is that the difference in gravity, Δg, above and below the cylinder is given by $\Delta g = -4\pi G\rho\Delta z$, where Δz is the vertical distance between the gravity measurements. This mathematical result means that borehole gravity is directly proportional to the lateral density variations of the Earth as a function of the depth of the borehole. The gravity signal represents vertical density variations integrated over a large horizontal extent. This is complimentary with other density measurements (such as detecting natural gamma radiation) that tend to sample material very close to the borehole. Gravity-derived density measurements provide a very nice picture (easily interpretable) of the density in the Earth surrounding the borehole.

Borehole gravity presents many difficult challenges for the gravity meter. First, the boreholes tend to have a small diameter, much smaller than most commonly made spring gravity sensors. **Figure 19** shows a photograph of the Gravilog quartz spring gravity borehole gravity sensor built by Scintrex Ltd. (Nind et al., 2007) next to a golf ball. The gravity sensor has a diameter of about 38.5 mm (1.5 in.). The ball in the figure contains a small quartz spring with electronic force feedback (capacitive), two layers of temperature control, two levels, and a leveling mechanism that permits the sensor to be made vertical. The high temperature of the borehole environment is also daunting. Scintrex Ltd. and Micro-g LaCoste Inc. are jointly developing spring borehole devices that operate in pressures up to 15 000 PSI and temperatures of 160 °C.

Borehole gravity is useful for exploration and for answering many different reservoir management problems. There are many applications in the oil and gas industry (Chapin and Ander, 1999) that range from determining the position and shape of salt domes to sequestration of gas and even managing water injection for enhanced production and many others. Borehole gravity can also be useful for civil engineering applications. For example, recently, borehole gravity was used to help answer questions about the geology underneath an underground nuclear waste treatment and immobilization plant in Hanford, Washington (MacQueen and Mann, 2007).

Figure 18 A10 deployed in Prudhoe Bay, from Micro-g LaCoste, Inc.

Figure 19 Quartz spring borehole gravity meter (Photo compliments of Scintrex, Ltd.).

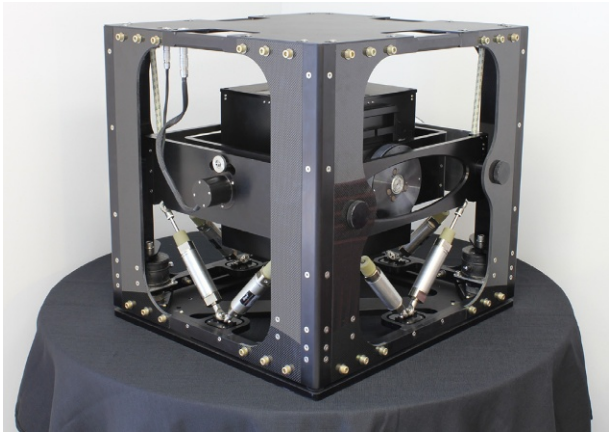


Figure 20 TAGS-6 airborne gravity meter.

Another challenging environment for gravity measurements is on dynamic platforms such as ships for marine applications or on aircraft (fixed wing and helicopters) for airborne applications. Dynamic environments are particularly difficult for gravity measurements because Einstein's equivalence principle states that it is impossible to distinguish between acceleration and gravity. Marine applications have the advantage that if one averages over periods longer than the wave motion, mean sea level is a stable surface. Airborne gravity is even more difficult because aircraft accelerations exist at all frequencies. Platform accelerations change not only the magnitude of gravity but also the direction of the gravity vector. Airborne gravity necessarily requires an independent means of determining aircraft accelerations so they can be subtracted from the measured gravity signal. The best way to do this is to use GPS to determine the motion of the aircraft in a differential mode. This requires a GPS antenna located on the aircraft (near the gravity sensor) and also one located at a fixed location on the ground. The difference in these two GPS signals is differentiated twice to find the acceleration of the aircraft in all three spatial directions. This can be subtracted from the measured value of gravity to determine the static gravity field emanating from the Earth below. **Figure 20** is a photograph of an airborne gravity system called TAGS-6. The gravity meter is housed inside of a platform that is servo controlled to remain vertical while the aircraft is executing its flight.

Airborne gravity is an excellent tool for collecting a large amount of gravity data over a large area. In some cases such as coastal areas or heavy jungle, there is no other feasible way to collect gravity data. In cases where land gravity measurements are possible, there is a trade-off between much higher data rates and ease of data collection and inherently lower precision (airborne gravity has about 10–50 times higher noise) and the reduced signal size (due to being farther from the density anomalies inside the Earth). There are other noncommercial applications such as geodesy where airborne gravity is very useful. Currently, the US government has a program called GRAV-D to provide a 2 cm gravity-based vertical datum across the country. According to Smith and Roman (2010), "Accurate heights are critical to many scientific endeavors, but particularly to understanding and protecting low-lying coastal ecosystems."

References

Block B and Moore RD (1966) Measurements in the earth mode frequency range by an electrostatic sensing and feedback gravimeter. *Journal of Geophysical Research* 71(18): 4361–4375.

Boy J-P, Llubes M, Hinderer J, and Florsch N (2003) A comparison of tidal ocean loading models using superconducting gravimeter data. *Journal of Geophysical Research* 108: 2193.

Brady JL, Ferguson JF, Hare JL, et al. (2008) Results of the world's first 4D microgravity surveillance of a waterflood – Prudhoe Bay, Alaska. *SPE Reservoir Evaluation & Engineering* 11(5): 824–831.

Chapin DA and Ander ME (1999) Borehole gravity: New life for borehole gravity. *AAPG Explorer* 20: 24–29.

Chartier JM, Labot J, Sasagawa G, Niebauer TM, and Hollander W (1993) A portable iodine stabilized He–Ne laser and its use in an absolute gravimeter. *IEEE Transactions on Instrumentation and Measurement* 42: 420–422.

Chris H, Lucien J, and LaCoste B (1996) Portrait of a scientist–inventor. *Earth in Space* 8(9): 12–13.

Davis R (2010) Fundamental constants: Big G revisited. *Nature* 468(7321): 181–183. http://dx.doi.org/10.1038/468181b.

Edwards A, Maki J, and Peterson D (1997) Gravity gradiometry as a tool for underground facility detection. *Journal of Environmental and Engineering Geophysics* 2.

Edwards AJ, Maki JT, Peterson DG, et al. (2011) Gravity probe B: Final results of a space experiment to test general relativity. *Physical Review Letters* 106: 221101.

Egorov KN and Martsinyak AI (1960) Determining the acceleration due to gravity at the location of the VNIIM. *Measurement Techniques* 3(8): 656–657.

Everitt CWF, DeBra DB, and Parkinson BW (1997) Gravity gradiometry as a tool for underground facility detection. *Journal of Environmental and Engineering Geophysics* 2: 139–143.

Everitt CWF, DeBra DB, and Parkinson BW (2011) Gravity probe B: Final results of a space experiment to test general relativity. *Physical Review Letters* 106.

Ferguson JF, Klopping FJ, Chen T, Seibert JE, Hare JL, and Brady JL (2008) The 4D microgravity method for waterflood surveillance: Part 3–4D absolute microgravity surveys at Prudhoe Bay, Alaska. *Geophysics* 73.

Francis O, Niebauer TM, Sasagawa G, Klopping F, and Gschwind J (1998) Calibration of a superconducting gravimeter by comparison with an absolute gravimeter FG5 in Boulder. *Geophysical Research Letters* 25(7): 1075–1078.

Francis O and van Dam TM (2003) Processing of the absolute data of the ICAG01. *Cahiers du Centre Européen de Géodynamique et de Séismologie* 22: 45–48.

Francis O and van Dam T (2006) Analysis of results of the international comparison of absolute gravimeters in Walferdange (Luxembourg) of November 2003. *Cahiers du Centre Europeen de Geodynamique et de Seismologie* 26: 1–24.

Giacomo P (1984) The new definition of the meter. *American Journal of Physics* 52(7): 607–613.

Hafner R and Widmer-Schnidrig R (2013) Signature of 3-D density structure in spectra of the spheroidal free oscillation 0S2. *Geophysical Journal International* 192: 285–294.

Heiskanen WA and Vening Meinesz FA (1958) *The Earth and its Gravity Field*. New York: McGraw-Hill.

Hugill A (1990) The Scintrex CG-3M autograv automated gravity meter, description and field results. In: *SEG Conference, San Francisco*.

Hugill AL and Valliant HD (1986) Limitations to the application of electrostatic feedback in gravity meters. *Journal of Geophysical Research* 91(B8): 8387–8392.

Imanishi Y, Higashi T, and Fukuda Y (2002) Calibration of the superconducting gravimeter T011 by parallel observation with the absolute gravimeter FG5 #210 – A Bayesian approach. *Geophysical Journal International* 151(3): 867–878.

Jacob T, Wahr J, Tad Pfeffer W, and Swenson S (2012) Recent contributions of glaciers and ice caps to sea level rise. *Nature* 482: 514–518.

Kasevich M and Chu S (1991) Atomic interferometry using stimulated Raman transitions. *Physical Review Letters* 67: 181–184.

Kater H (1818) *Philosophical Transactions of the Royal Society of London* 108: 33.

Kennard EH (1938) *Kinetic Theory of Gases*. New York: McGraw-Hill.

Kennedy J, Creutzfeldt B, and Ferre PA (2012) Monitoring vadose zone infiltration with time-lapse gravity data at a municipal recharge and withdrawal facility (invited). In: *GSA Annual Meeting, November 4–7, Charlotte, NC*.

Lambert A, Courtier N, and James TS (2006) Long-term monitoring by absolute gravimetry: Tides to postglacial rebound. *Journal of Geodynamics* 41(1–3): 307–317 (ESS Cont.# 2005056).

Lambert A, Courtier N, Sasagawa GS, et al. (2001) New constraints on Laurentide postglacial rebound from absolute gravity measurements. *Geophysical Research Letters* 28(10): 2109–2112.

Lambert A, James TS, and Thorleifson LH (1998) Combining geomorphological and geodetic data to determine postglacial tilting in Manitoba. *Journal of Paleolimnology* 19(3): 365–376.
Lange B (1964) The Drag-free satellite. *AIAA Journal* 2(9): 1590–1606.
Louchet-Chauvet A, Farah T, Bodart Q, et al. (2011) The influence of transverse motion within an atomic gravimeter. *New Journal of Physics* 13: 065025.
MacQueen JD and Mann E (2007) *Borehole Gravity Meter Surveys at the Waste Treatment Plant, Hanford, Washington*. Richland, WA: Pacific Northwest Laboratories, US Government Publication, PNNL-16490. http://www.pnl.gov/main/publications/external/technical_reports/PNNL-16490.pdf.
McGarva Bruckshaw J (1941) Gravity meters. *Proceedings of the Physical Society* 53: 449–467.
Moore RD and Farrell WE (1970) Linearization and calibration of electrostatically feedback gravity meters. *Journal of Geophysical Research* 75(5): 928–932.
Newell DB, Steiner RL, and Williams ER (1998) An accurate measurement of Planck's constant. *Physical Review Letters* 81: 2404–2407.
Newell DB, Steiner RL, and Williams ER (1999) *The NIST Watt Balance: Recent Results and Future Plans*. Atlanta, GE: American Physical Society.
Niebauer T (1987) *New Absolute Gravity Instruments for Physics and Geophysics*. Thesis in Physics, Boulder, CO: University of Colorado.
Niebauer TM, Billson R, Ellis B, Mason B, van Westrum D, and Klopping F (2011) Simultaneous gravity and gradient measurements from a recoil-compensated absolute gravimeter. *Metrologia* 48(3): 154.
Niebauer TM, MacQueen J, Aliod D, and Francis O (2011) Monitoring earthquakes with gravity meters. *Geodesy and Geodynamics* 2(3): 71–75.
Niebauer TM, Sasagawa GS, Faller JE, Hilt R, and Klopping F (1995) A new generation of absolute gravimeters. *Metrologia* 32(3): 346–352.
Niebauer TM, Schiel A, and van Westrum D (2006) Complex heterodyne for undersampled chirped sinusoidal signals. *Applied Optics* 45: 8322–8330.
Nind C, Seigel HO, Chouteau M, and Giroux B (2007) Development of a borehole gravimeter for mining applications. *First Break* 25(7).
Park J, Song TR, Tromp J, et al. (2005) Long-period behavior of the 26 December 2004 Sumatra–Andaman earthquake from its excitation of Earth's free oscillations. *Science* 308: 1139–1144.
Parker PR, Zumberge MA, and Parker RL (1995) A new method for fringe-signal processing in absolute gravity meters. *Manuscripta Geodaetica* 20(3): 173–181.
Prothero WA (1967) *A Cryogenic Gravimeter*. PhD Thesis, San Diego: University of California.
Prothero WA and Goodkind JM (1968) A superconducting gravimeter. *Review of Scientific Instruments* 39: 1257–1262.
Quinn TJ (2002) Measuring big G. *Nature* 408: 919–921.
Raab F and Coyne D (1997) Effect of microseismic noise on a LIGO interferometer, LIGO-T960187-01-D. http://www.ligo-wa.caltech.edu/ligo_science/museism.pdf.
Richter B, Wilmes H, and Nowak I (1995) The Frankfurt calibration system for relative gravimeters. *Metrologia* 32(3): 217.
Rinker R (1980) Phd Thesis, University of Colorado.
Sakuma A (1970) Recent developments in the absolute measurement of gravitational acceleration. *National Bureau of Standards* 343: 447–456, Special Publication.
Sasagawa G, Klopping FJ, Niebauer TM, Faller JE, and Hilt R (1995) Intracomparison tests of the FG5 absolute gravity meters. *Geophysical Research Letter* 22: 461–464.
Seigel HO, Brcic I, and Mistry P (1993) A high precision, μGal resolution, land gravimeter with worldwide range. In: Seigel HO (ed.) *A Guide to High Precision Land Gravimeter Surveys*. Concord, ON: Scintrex Ltd.
Smith DA and Roman DR (2010) How NOAA's GRAV-D project impacts and contributes to NOAA science. In: *GRAV-D PI, NOAA's National Geodetic Survey, April 29, 2010* .
Tate DR (1968) Acceleration due to gravity at the National Bureau of Standards. *Journal of Research of the National Bureau of Standards C, Engineering and Instrumentation* 72C(1), National Bureau of Standards Monograph 107.
Torge W (2001) *Geodesy*. KG, Berlin, Germany: Walter de Gruyter GmbH & Co.
Valliant HD (1991) Gravity meter calibration at LaCoste and Romberg. *Geophysics* 56: 705–711.
Valliant HD, Gagnon C, and Halpenny JF (1986) An inherently linear electrostatic feedback method for gravity meters. *Journal of Geophysical Research* 91: 10463–10469.
Van Camp M and Francis O (2007) Is the instrumental drift of superconducting gravimeters a linear or exponential function of time? *Journal of Geodesy* 81: 337–344.
Van Camp M, Williams SDP, and Francis O (2005) Uncertainty of absolute gravity measurements. *Journal of Geophysical Research* 110: B05406. http://dx.doi.org/10.1029/2004JB003497.
Vitushkin L, Becker M, Jiang Z, et al. (2002) Results of the sixth international comparison of absolute gravimeters ICAG-2001. *Metrologia* 39(5): 407–427.
Weiss R and Block B (1965) A gravimeter to monitor the ${}_0S_0$ dilational mode of the earth. *Journal of Geophysical Research* 70(22).
Zumberge MA (1981) *Internal Working note for California Institute of Technology LIGO Project-MS 51–53*. Boulder, CO: University of Colorado.

This page intentionally left blank

3.04 Superconducting Gravimetry

J Hinderer, Institut de Physique du Globe de Strasbourg, Strasbourg, France
D Crossley, St Louis University, Saint Louis, MO, USA
RJ Warburton, GWR Instruments, San Diego, CA, USA

© 2015 Elsevier B.V. All rights reserved.

3.04.1 The Superconducting Gravimeter 60
3.04.1.1 Historical 60
3.04.1.1.1 Early years at the University of California at San Diego 60
3.04.1.1.2 Early commercial model TT instruments (1981–94) 60
3.04.1.1.3 The compact SG (1994–2002) 61
3.04.1.1.4 The observatory SG 62
3.04.1.1.5 Development of the dual-sphere design 63
3.04.1.1.6 *i*Grav® superconducting gravimeter 64
3.04.1.2 Basic Principles of Operation 65
3.04.1.2.1 Superconducting components 65
3.04.1.2.2 Displacement transducer and feedback system 67
3.04.1.2.3 Temperature control 67
3.04.1.2.4 Tilt compensation system 67
3.04.1.2.5 Sphere and sphere resonance 68
3.04.1.3 Instrument Performance 68
3.04.1.3.1 Instrument drift 68
3.04.1.3.2 Calibration stability 68
3.04.1.3.3 Instrumental noise and precision 68
3.04.1.4 User Requirements 70
3.04.1.4.1 Operation and maintenance 70
3.04.1.4.2 Site location 70
3.04.1.4.3 Site noise 70
3.04.1.4.4 Site stability 71
3.04.2 SG Data Analysis 71
3.04.2.1 Preprocessing 71
3.04.2.1.1 Second data sampled from an SG 71
3.04.2.1.2 Minute data from the GGP database (ICET) 71
3.04.2.1.3 Remove–restore technique 74
3.04.2.1.4 Treatment of gaps in gravity data 74
3.04.2.1.5 Disturbances and offsets 74
3.04.2.1.6 Automatic procedures 75
3.04.2.1.7 Processing for different purposes 75
3.04.2.1.8 Restoring the signal 77
3.04.2.2 Solid Earth and Ocean Tides 77
3.04.2.2.1 Tide-generating potential 77
3.04.2.2.2 Elastic response of the Earth 78
3.04.2.2.3 Ocean tides and loading 79
3.04.2.2.4 Tidal analysis 79
3.04.2.3 Atmospheric Pressure Effects 80
3.04.2.3.1 Single admittance factors 80
3.04.2.3.2 Frequency-dependent admittance 80
3.04.2.3.3 Green's functions and nonlocal pressure corrections 81
3.04.2.3.4 Three-dimensional atmospheric corrections 82
3.04.2.4 Calibration Issues 83
3.04.2.4.1 Basics 83
3.04.2.4.2 Amplitude calibration and relative methods 84
3.04.2.4.3 Amplitude calibration and absolute methods 85
3.04.2.4.4 Phase calibration 85
3.04.2.5 Other Corrections to Residual Gravity 86
3.04.2.5.1 Polar motion 86
3.04.2.5.2 Instrument drift 86

3.04.2.5.3 Hydrology 87
3.04.2.5.4 Residual gravity 87
3.04.3 Scientific Achievements Using SGs 87
3.04.3.1 The Global Geodynamics Project 87
3.04.3.2 Seismic and Subseismic Signals 89
3.04.3.3 Atmospheric Loading 92
3.04.3.4 Tides and Nearly Diurnal Earth Wobbles 94
3.04.3.4.1 Resonance effects in diurnal tides 94
3.04.3.4.2 Ocean loading 95
3.04.3.5 Nontidal Ocean Circulation 96
3.04.3.6 Hydrology 97
3.04.3.7 Ice Melting 101
3.04.3.8 Earth Rotation 102
3.04.3.9 Tectonic Effects 102
3.04.3.10 Ground/Satellite Gravity Field Comparison 104
3.04.3.11 Future Possibilities 106
3.04.3.12 Conclusions 107
Acknowledgments 108
References 108

3.04.1 The Superconducting Gravimeter

3.04.1.1 Historical

3.04.1.1.1 Early years at the University of California at San Diego

The superconducting gravimeter (SG) was first introduced by Prothero and Goodkind (1968) as part of Prothero's (1967) thesis work on the design and development of the instrument, undertaken at the University of California at San Diego (UCSD). The SG broke new ground in geophysics instrumentation and was an elegant realization of the principles of superconductivity. Although the basic sensor configuration has remained unchanged for nearly 40 years, continuous improvements in all other aspects of the original design have successfully converted the SG from a prototype laboratory instrument to a reliable research tool. (Note that in this chapter, we frequently use traditional gravity abbreviations: 1 microgal = 1 μGal = 10 nm s^{-2}, 1 nanogal = 1 nGal = 0.01 nm s^{-2}, and cpd = cycles per (solar) day.)

In 1970, Richard Warburton became a postdoctoral student with John Goodkind, a collaboration that was the foundation for the eventual line of commercial SGs. Richard Reineman, an undergraduate laboratory assistant working with William Prothero in 1969, was integral to the effort as a development technician with Goodkind and Warburton. Prothero and Goodkind (1972) published the first observations taken over a 4-month period and obtained precise tidal amplitude and phases, new information on ocean tide loading, and a recording of seismic normal modes following the 7.1 magnitude Kamchatka earthquake in 1969. Within a few years, the UCSD group generated significant papers using SG data on ocean tide loading (Warburton et al., 1975) and the effects of barometric pressure on gravity (Warburton and Goodkind, 1977). This phase of the SG research culminated with a detailed tidal analysis of 18 months of data that included the first estimate of the effect of the nearly diurnal wobble on the resonant amplification of small diurnal tidal phases (Warburton and Goodkind, 1978). These papers are still a recommended reading for those interested in the basic issues concerning the treatment of gravity data.

3.04.1.1.2 Early commercial model TT instruments (1981–94)

Early publications, as well as presentations at various conferences, caught the attention of Paul Melchior (Royal Observatory of Brussels, Belgium (ROB)) and Rudolf Brein and Bernd Richter (Institut für Angewandte Geodäsie, Germany (IfAG), now known as Bundesamt für Kartographie und Geodäsie (BKG)), who contacted Goodkind about the possibility of using SGs to expand their previous work based on LaCoste–Romberg (LCR) gravity meters. As a result, the commercial venture GWR Instruments (Goodkind, Warburton, and Reineman) were formed in 1979 to manufacture two SGs, one for ROB and one for IfAG. From this point on, two different design streams continued: Goodkind refined the original UCSD instruments and used them to develop new areas of geophysical research and GWR began the manufacture of instruments for other scientific groups from their San Diego facilities. Eric Brinton joined GWR in 1986 to continue the development of refrigeration, electronics, and data acquisition systems.

Melchior purchased a GWR model TT30 dewar similar to those used at UCSD. The SG sensor was cooled by simple insertion through the 5 in. diameter neck of a dewar with its 200 l internal volume (the 'belly') filled with liquid helium. In a typical dewar, the belly is surrounded by a vacuum space, which contains two radiation shields with many thin layers of aluminized Mylar ('superinsulation') placed on the surface of the belly and shields. Hold time critically depends on efficiently using the cooling power of the gas (enthalpy) as it flows past the shields and neck, before exhausting at room temperature. The dewar was suspended from a large concrete pier using a sturdy isosceles metal frame with three contact points to the pier – two adjustable micrometer screws and the third fixed point. The dewar was leveled by adjusting the height of the micrometer screws with respect to the fixed rear point.

In 1981, the model TT30 was installed in the basement vault at ROB. Visiting scientists who were familiar with modern SG installations would have been greeted by an eerie silence – there was no compressor noise, not even in an adjacent room.

Silence had its disadvantage, however, as almost 200 l of liquid helium had to be replenished every 3 weeks or so and each of these refills caused unpleasant disturbances to the data stream. Despite many problems that originated from a helium leak in the TT30 vacuum can lid (described in detail in De Meyer and Ducarme, 1989), this instrument was to continue without major interruptions for nearly 18 years until it was decommissioned in 2000. Early tidal results from the Brussels SG can be found in Ducarme (1983).

Helium was very expensive in Germany, so IfAG/BKG asked GWR to develop a refrigerated dewar system for their instrument. These systems use a cryogenic refrigerator (cold head and compressor) to intercept and reduce the flow of heat via radiation and conduction from the outside of the dewar to its belly. This reduces the rate of boil-off, extending the 'hold time' of the liquid helium. Hold time depends on the cooling power of the refrigerator being used and how well it can be thermally coupled to the neck and radiation shields. On the first TT40 refrigerated dewar, the cold head was bolted onto the top of the dewar wall and penetrated into the vacuum space through a special port sealed with an O-ring. The cold head's two cooling stages (at 65 and 11 K) were connected directly to the inner and outer radiation shields using copper braid. The TT40 design was extremely successful, with a hold time of well over 400 versus 50 days unrefrigerated. This project was the beginning of a collaboration between GWR Instruments and IfAG/BKG to develop new and improved SG models, which continued for many years.

Access to a commercial instrument proved to be a landmark opportunity for the geodetic and gravity community. In 1981, SG model TT40 was installed in the basement of a castle at Bad Homburg, near Frankfurt, Germany. At the IUGG (International Union of Geodesy and Geophysics) meeting in Hamburg, Richter (1983) presented a paper on data from the TT40 that showed the first gravitational determination of the 14-month Chandler component of polar motion, with an amplitude of ± 5 μGal. This took the audience by surprise, to say the least, and convincingly demonstrated the capabilities of the new gravimeter.

At the same time, Goodkind (1983) repeated his determination of the nearly diurnal wobble parameters from the tidal amplitudes using the data from 1978, but the problem of accurately computing the ocean tidal loading of small waves remained a limiting factor. A few years later, Richter (1985) reported on the extension of his data set to 3 years of successful SG operation.

In 1985, Richter installed a second instrument – a new model TT60 – at Bad Homburg, and until the beginning of 1987, he obtained parallel recording with the original TT40. His thesis (Richter, 1987) contained many interesting insights into the operation of the instrument and its data, but, being in German, it was not widely read. A significant result was that the gravity residuals from both instruments were highly correlated at the sub-μGal level (Richter, 1990), indicating that significant geophysical signals were still left in the data at this level. At the time, the cause was ascribed to the atmosphere, but this was before the environmental influence of hydrology on gravity became widely appreciated. Following the experimental work of Richter and the theoretical speculations of Melchior and Ducarme (1986), new refrigerated model TT70s were installed in Wuhan, China, in 1986 (Hsu et al., 1989) and in Strasbourg, France, in 1987 (Hinderer and Legros, 1989). The model TT70 inaugurated the use of internal tiltmeters and thermal levelers to automatically keep the SG aligned.

As with the TT40 and TT60, the first two TT70s were manufactured with the cold head bolted into the top of the dewar and penetrating the vacuum space. Although very efficient, this design made it difficult to service the cold head without warming the dewar to room temperature. Later, TT70 models were manufactured with the cold head supported by a separate frame and inserted through the neck of the dewar. Cooling power was coupled to the neck and radiation shields solely via helium gas. In the new TT70 design, the cold head could easily be removed for servicing or removal of any ice that may build up between the cold head and the gravity sensor unit (Warburton and Brinton, 1995). New TT70s were soon installed at the National Astronomical Observatory of Japan in Mizusawa in 1988, and two were located side by side at Kyoto University in the same year (Tsubokawa, 1991).

Meanwhile, in 1989, Richter moved the TT60 from Bad Homburg to Wettzell, a fiducial station of the German geodetic network, and a TT70 was installed at Cantley, Canada, in the same year (Bower et al., 1991). Approximately 12 TT70 SGs were manufactured from 1986 to 1994, and 6 of these instruments are still operating as part of the Global Geodynamics Project (GGP) network. TT70 dewars are 150 cm tall, have a 80 cm diameter, and weigh 150 kg. They require an annual transfer of 200 l of liquid helium and cold head service at 1- to 2-year intervals. In 1993, it was found that the TT70 SG was less susceptible to horizontal noise sources when the dewar was mounted from the bottom (Warburton and Brinton, 1995). **Figure 1** shows the TT70 operating at Cantley, Canada, after modification to the bottom-mounted configuration. The large concrete pier, previously used to support the dewar, now only supports the cold head.

3.04.1.1.3 The compact SG (1994–2002)

In 1993, GWR produced the much smaller 125 l compact dewar designed to operate on a 1 m^2 pier or platform, so that

Figure 1 An example of the TT70-type instrument installed at Cantley, Canada, in 1989. The rack on the far left contains the chart recorders and noncritical electronics; the more sensitive components are in the enclosed temperature controlled rack next to it. The 200 l dewar sits on 3 feet placed on small granite blocks. The front 2 feet with the X and Y thermal levelers attached to the dewar bottom are visible. The cold head is supported from a frame resting on the top of a concrete pier. Normally, the entire SG is enclosed by thermal insulation.

Figure 2 A compact instrument C023 installed at Medicina, Italy.

it could be easily operated at many preexisting geodetic installations. The compact SG is 104 cm high and 66 cm wide and weighs 90 kg. The SG sensor is built into the dewar belly, which allows the neck and radiation shields to be custom-designed to mirror the dimensions of the cold head. The compact dewar uses the same APD Cryogenics DE202 cold head and HC-2 helium compressor as used in models TT60 and TT70. However, the smaller neck diameter and volume dramatically reduce the heat load on the outer radiation shield, and the improved neck–cold head interface more efficiently uses the cold head cooling power. As a result, the dewar efficiency is doubled and < 100 l of liquid helium is used annually.

Figure 2 shows compact C023 SG operating in Medicina, Italy (Romagnoli et al., 2003). The APD HC-2 compressor and its water chiller are on the left side (rear), with flexible stainless hoses connecting the compressed helium gas to the cold head. The two thermal levelers and a third fixed point are attached to an aluminum band wrapped around the circumference of the dewar. These are supported by 3 feet that rest on small granite blocks placed upon the floor. The cold head is supported and centered in the neck using a metal tripod support frame. The neck–cold head interface is sealed with a rubber gasket, which prevents air from entering and provides vibration isolation between the cold head and the SG sensor. The addition of stiff internal spokes placed between the inner belly and the outer dewar wall makes the compact dewar less sensitive to horizontal noise. This structural change produces lower noise levels than observed in previous SGs (Boy et al., 2000).

The first compact SG C021 was tested next to T002 at the Royal Observatory of Brussels before being moved in 1995 to a seismic station in Membach, Belgium, where it is still operating today. The SG is installed in a separate room at the end of a 100 m long tunnel. Due to its proximity to long-period seismometers, care was taken to minimize and measure vibrations produced by the cryogenic refrigeration system (Van Camp, 1995). Eleven compact SGs (including two with dual sensors) were manufactured from 1994 to 2002 and about nine are still operating at GGP stations.

3.04.1.1.4 The observatory SG

3.04.1.1.4.1 Ultralong-hold time dewar

Although compact dewars were extremely successful, the goal remained to further reduce helium consumption and annual disturbances from helium transfers. By 1997, the first commercial (and somewhat) practical 4 K refrigeration systems became available. The Leybold Vacuum product KelKool 4.2 GM cold head produced cooling power of 50 W at 50 K at its upper stage and 0.5 W at 4 K at its lower stage. It cooled well below the 4.2 K liquefaction temperature of helium and produced more than five times the cooling power of the APD DE202. Soon afterward, GWR produced an ultralong-hold time dewar (ULHD) based on the 125 l compact dewar design, with its neck modified to accommodate the much larger 4.2 GM cold head. This creates a closed-cycle system, because the helium gas condenses in the neck on the lower stage and drips back into the storage volume of the dewar (Richter and Warburton, 1998). This success pointed to the future in which SGs could operate indefinitely without transferring liquid helium or consuming liquid helium. The first ULHD system is shown in Figure 3.

Figure 3 One of the first dual-sphere instruments (CD029, now at Wettzell, Germany) with an ultralong-hold time dewar using a Leybold KelKool 4.2 K GM cold head. The support cranes used to insert and remove the cold head are shown in the background.

Two dual-sensor SGs using the KelKool 4.2 GM coldhead (CD029 and CD030) were installed at Wettzell and at Bad Homburg, Germany, in June 1999. The cold heads were extremely reliable and only required maintenance only at $\sim$3-year intervals. However, the cold heads were too heavy for one person to handle and required a support crane for insertion and removal. Additionally, the combination of compressor and water chiller required 7 kW of power versus 2 kW power for the DE202 system. Two more ULHD systems were manufactured: R038, which was installed in Concepción,

Chile, in December 2002, and C043, which replaced T016 at Syowa Station, Antarctica, in March 2003. Neither of these stations could supply 7 kW, so these ULHD were designed to use a Leybold 4.2 Lab cold head, which with water cooling reduced the power load to 3.5 kW (Warburton et al., 2000). After several years of operation, eventually, all four of these Leybold-based systems were replaced with Sumitomo SHI RDK-101 cold heads, which are discussed in the next section.

3.04.1.1.4.2 Observatory dewar

Within a year of shipping the C043, Leybold stopped manufacturing the 4.2 Lab cold head, threatening the 5-year effort to develop a closed-cycle system. Fortunately, Sumitomo Heavy Industries, which had extensive experience with large 4 K cryocoolers, entered the market in the same year with a new smaller refrigeration system, the SHI RDK-101 cold head and CNA-11 compressor. The RDK-101 cold head is about the same physical size as the APD 202, but it uses small 16 mm diameter flex hoses and is easily handled by one person. The CNA-11 compressor only uses 1.3 kW power and is air-cooled. As a result of the smaller size and power, the RDK-101 has less than half the cooling and liquefaction power of the Leybold 4.2 Lab. Since its introduction, the RDK-101 cold head has proved to be a very reliable replacement for both the DE202 and the Leybold 4.2 Lab cold heads.

Over the next 2 years, GWR reengineered and tested several cold head–dewar interfaces to fully utilize the entirety of the RDK-101's cooling power. Small dewars have several attractive features: They are visually appealing, lighter, and easier to move and install, and most importantly, the input heat load decreases with surface area. Larger dewars provide longer hold times in the event of cold head failure. After experimenting with dewar volumes as small as 10 l, GWR chose a 35 l capacity for its observatory dewar design. This compromise allows 20 days of operation in failure mode and enough excess cooling capacity to liquefy helium gas at a rate higher than 1 l day^{-1}.

The first OSG-O040 was installed in Walferdange, Luxembourg, in December 2003; the second in S. Korea in March 2005; and two more installed in Taiwan in March 2006. Nineteen OSG SGs were manufactured between 2003 and 2012, and about 14 are presently operating at GGP sites. Figure 4 shows the OSG-O049 dewar system, with all its control electronics and data acquisition system.

3.04.1.1.4.3 Data acquisition system and remote control

In the early years, GWR supplied analog electronics (Gravimeter Electronics Package (GEP)) for controlling the gravity, temperature, and tilt functions; and a current supply (Dual Power Supply and Heater Pulser (DPS)) for sphere levitation. Analog filters were copies of those used for the International Deployment of Accelerometers (IDA) network of LCR gravity meters and took the approach of dividing the signal between a low-passed tidal gravity output – the main system output – and a short-period high-passed output that displayed signals such as earthquakes and disturbances. Each user provided their own data acquisition system to record the tidal gravity output and the atmospheric pressure. The high-frequency signal was only recorded on a strip chart recorder for visual observation.

Figure 4 The 'observatory model' SG, SHI RDK-101 cold head, and CAN-11 compressor. All of the control and data acquisition electronics are contained inside one temperature-regulated electronics enclosure. During installation, the front panel is removed and all the controls are accessed by the local keyboard and computer screen normally stored inside the enclosure. After installation, all functions and data retrieval are remotely accessed through an Internet connection. The dewar is 42 cm in diameter and the combined height of the dewar and cold head is 130 cm.

As more users began to acquire SGs for a variety of purposes, it became a standard practice to sample at a high rate (1–10 s) the full signal (both tides and seismic frequencies). Nevertheless, it took some time for the gravimeter community to develop a common set of standards for the acquisition and exchange of SG data. To help this process along, GWR provided a replacement gravity card (GGP gravity card) with a filter designed for 1 s sampling and additional circuitry for measuring the phase response of the gravimeter system (GGP Newsletters #2 and #3; http://www.eas.slu.edu/GGP/ggpnews.html).

The first GWR data acquisition system, manufactured in 1995, used CSGI software running on a PC with a QNX operating system. Soon, however, the uncertain future of the QNX operating system convinced GWR to develop a Windows-based system. In 1999, GWR and BKG reported an ambitious project to control the SG remotely (Warburton et al., 2000). The prototype remote SG R038 was installed at Concepción, Chile, in December 2002, and it met almost all of GWR's goals for remote operation. In 2004, GWR decided that all new observatory SGs should be provided with a GWR data acquisition system with remote-control capabilities. This decision was made not only to further standardize GGP data but also to enable GWR to remotely diagnose problems as they arose and solve them without traveling to the site of operation.

3.04.1.1.5 Development of the dual-sphere design

In the early commercial SGs manufactured up to 1990, offsets (or 'tares') occurred in the SG gravity records that affected both

long-term stability and the measurement of tidal factors (Harnisch and Harnisch, 1995; Hinderer et al., 1994; Seama et al., 1993). The offsets could be quite large (up to 100 μGal) if caused by power failures or mechanical shocks from earthquakes or liquid helium transfer. Small instrumental offsets <5 μGal could also occur at random intervals not associated with outside disturbances. Though rapid offsets larger than 0.2 μGal and <1 min could easily be detected and corrected (Merriam et al., 2001), there was concern that the residual data would depend arbitrarily on the threshold value chosen in automatic offset-detection programs (Harnisch and Harnisch, 1997). When two SGs were operated side by side, the difference in recordings provided a much clearer determination of the occurrence of offsets (Klopping et al., 1995).

A dual-sphere SG was manufactured to solve the instrument offset problem (Richter and Warburton, 1998), with the knowledge that random offsets will seldom occur in two sensors simultaneously. The two spheres are mounted one above the other and separated by about 20 cm. The lower sensor is manufactured exactly like the previous single-sphere sensors, and the temperature and tilt control remain the same. The upper sensor is manufactured with two pairs of side coils, which are used to apply adjustable horizontal forces to the sphere. Small differences in the sphere masses, superconducting shield, coil windings, and machining tolerances produce magnetic asymmetries that are not identical in the two sensors. These asymmetries produce slightly different tilt minima in the two sensors. Small currents are trapped in the side coils to align the tilt minimum of the upper sensor with the lower sensor.

The complications of a dual-sphere system are justified by providing a built-in instrumental offset detector. Because the outputs are treated as signals from two different gravimeters, the user can combine the processed data sets, select the least disturbed sphere output for any one time period, or convert the two signals into a gravity gradient by using the known vertical separation. CD029 was the first dual-sphere SG produced, and it was tested at Bad Homburg beginning July 1998, before being moved permanently to Wettzell in November 1998 (Harnisch et al., 2000). Results at Bad Homburg showed that careful calibration of both amplitude and phase was required to minimize the difference signal and that offsets of a few tenths of a μGal could indeed be easily detected. Surprisingly, after relocation to Wettzell, no spontaneous random offset exceeding about 0.1 μGal was observed. Larger offsets that did occur were caused by failures in the cooling system or during extensive maintenance procedures (cold head maintenance, He refills, or removal of accumulated ice from the neck).

More recent data from all four dual-sphere SGs support the early observations of CD029 (Kroner et al., 2004, 2005). No offsets have occurred in CD029 at Wettzell or D034 at Moxa, Germany, and only one or two offsets per year are observed in CD028 at Bad Homburg and in CD037 at Sutherland, South Africa. From these data, GWR concludes that the changes in its manufacturing process – in particular, improvements to the niobium spheres and shields – have greatly reduced the incidence of random offsets. Given such a success, one might argue that the dual-sphere SG is no longer needed. However, as gravity changes are examined with higher and higher resolution, they may still be used to discriminate sub-μGal observations of instrumental origin (see Figure 3 of Meurers, 2004) or illuminate more subtle instrumental effects that need improvement (Kroner et al., 2005). More importantly, the gradient signal itself may yet prove useful in modeling environmental gravity variations.

3.04.1.1.6 iGrav® superconducting gravimeter

Several factors – namely, the high cost, the complexity of equipment, and the perceived difficulty of operation and maintenance – have historically restricted the operation of SGs primarily to observatory buildings. These factors, in addition to the need for suitable infrastructure, have prevented SGs from use in many applications where their low drift and noise could make significant contributions. To encourage new environments of use, GWR designed the *i*Grav® superconducting gravimeter (*i*SG) not only to reduce the cost, size, and weight but also to make SGs much more portable and easy to use (cf. **Figure 5**). The most significant change is to the electronics – 90% of which now fit in the head of the dewar. In addition, the dewar volume is reduced from 35 to 16 l, its height is reduced from 1.14 to 0.82 m, and its weight is reduced by removing the heavy thermal levelers previously attached to the sides of the OSG dewar. Instead, smaller and lighter thermal levelers attach to a baseplate, which is set on the floor and stabilized with six leveling screws. The cold head frame and alignment mechanism is lighter and simpler to use with only four screws holding the cold head in place. Overall, the total weight of the *i*SG – excluding the refrigeration system – is reduced by a factor of 3.4 from the OSG, from 230 to 68 kg. The dewar requires a pier of only 0.8×0.8 m and the total system can now operate in a 2×0.8 m area.

The *i*SG gravity design uses the same principles as all other SGs; however, two changes simplify its operation. First, the magnetic force gradient (the 'spring constant') is permanently set at the factory by adjusting the ratio of the turns of the upper

Figure 5 *i*Grav® SG model (*i*SG) – dewar and cold head at the left, DAC and monitor on top of UPS in the middle, and CNA-11 compressor and 8500 l He gas cylinder at the right.

and lower magnet coils and wiring them in series. Only the magnitude of one current needs adjustment until the total levitation force precisely balances local g at the site of operation. Second, the sphere is centered in the capacitance displacement transducer by adjusting and trapping a small current in an independent centering coil. These two changes eliminate the difficult task – integral to the performance of earlier SGs – of adjusting the gradient and centering the sphere. Initial setup and continuing operation thus require less intensive user training. These and other improvements have markedly decreased initial drifts, which are largely absent after a few weeks of operation. However, longer records and precise comparisons to AGs are still required to determine whether the long-term drifts of *i*SGs are less than the 2–4 μGal year^{-1} rates observed in the best OSGs.

The *i*SG can be moved between locations with the sphere levitated, and after relocation, the user only needs to adjust the current in the centering coil to recenter the sphere. Most importantly, initial results show that there is no change in scale factor or drift after such relocation. This means that gravity changes in time ($dg(t)/dt$) can be measured immediately after the *i*SG is moved without the need to discriminate between instrumental drift and geophysical signals of interest.

Wilson et al. (2012) were the first to make the OSG transportable by operating it in two moveable field enclosures. Following their lead, GWR, in close cooperation with the GFZ German Research Centre for Geosciences, developed two NEMA-4 protective class enclosures for operating the *i*SG over a wide range of temperatures (−17 to 46 °C) and environmental conditions. **Figure 6** shows the first prototype operating at Tucson, Arizona (Creutzfeldt et al., 2012a,b; Kennedy et al., 2014). The enclosure in the foreground houses the *i*SG, dewar, electronics, cold head, and barometer. This unit is climate-controlled to keep the temperature within ±3 °C and has a removable top lid that allows access for cold head removal and maintenance. The second enclosure in the background houses the compressor, UPS, storage batteries, and He gas cylinders. Its climate-control design keeps the compressor operating within the manufacturer-recommended temperature specifications. A flexible hose containing the compressor hoses, power, and communication connects the two enclosures and allows a separation of up to 12 m. The design concept minimizes the impact of the enclosure (its 'rain shadow'), which can significantly alter the interpretation of gravity data in the context of hydrologic studies (Creutzfeldt et al., 2008, 2010b). The only disturbance to the site is from construction of a small concrete pier below the *i*SG enclosure.

Figure 6 GWR *i*SG enclosures operating in Tucson, Arizona. Figure courtesy of Jeffrey Kennedy.

This type of enclosure system works well where reliable AC power is available to operate the compressor (1.3 kW) and electronics (100 W). During a power failure, the UPS backs up the electronics, and the number of batteries can be chosen to accommodate power failures of 1–3 days, without loss of data. The compressor is not backed up, however, so the length of operation of the SG sensor relies upon the volume of helium stored in the dewar. The 16 l *i*SG dewar loses about 10% per day, so power failures of 1–3 days are easily covered. Data will be lost for power failures longer than 3 days, though the sensor will remain at 4 K and undisturbed for up to 10 days. When power is restored, the system recovers and data taking commences. Within 1 day, the valve to the He gas cylinder valve can be opened to restart the liquefaction process that will refill the dewar in preparation for the next power failure.

This small enclosure design cannot meet the needs of all users and field sites. Some users require more space for making absolute gravimeter (AG) measurements next to the SG or for operating other gravity meters or geophysical equipment. Some sites may experience long power failures, which exceed the hold time of the dewar, and very remote sites may have no power available at all. To operate under these conditions, GWR, in close cooperation with the Geological Survey of Japan/AIST (Sugihara and Nawa, 2012; Sugihara et al., 2013), built a larger enclosure that was first installed in a CO_2 sequestration field at Gordon Creek, Utah, and then moved to a CO_2/enhanced oil recovery field near Farnsworth, Texas.

3.04.1.2 Basic Principles of Operation

3.04.1.2.1 Superconducting components

Seismometers and relative gravimeters depend on a test mass that is suspended by a spring attached to the instrument support. A change in gravity or ground motion generates a voltage that becomes the output signal (velocity or acceleration). This system works well as a seismometer in many modified forms and is still used successfully in the LCR and Scintrex models of field gravimeters. But at periods longer than the seismic normal-mode range – for example, beginning at 4 h for tides – the mechanical aspects of spring suspension cause erratic drift, even in a thermally well-regulated environment, that is difficult to remove with postprocessing. Field gravimeters repeatedly occupy a reference station to monitor this drift, and observatory spring instruments have to be rezeroed when the signal exceeds the range of the voltmeter. Since the 1980s, spring gravimeters have incorporated electrostatic feedback that considerably improves their linearity and drift performance (e.g., Larson and Harrison, 1986).

The SG almost completely solves the drift problem by replacing the mechanical spring with the levitation of a test mass using a magnetic suspension. **Figure 7** shows a diagram of the gravity sensing unit; the three major superconducting elements

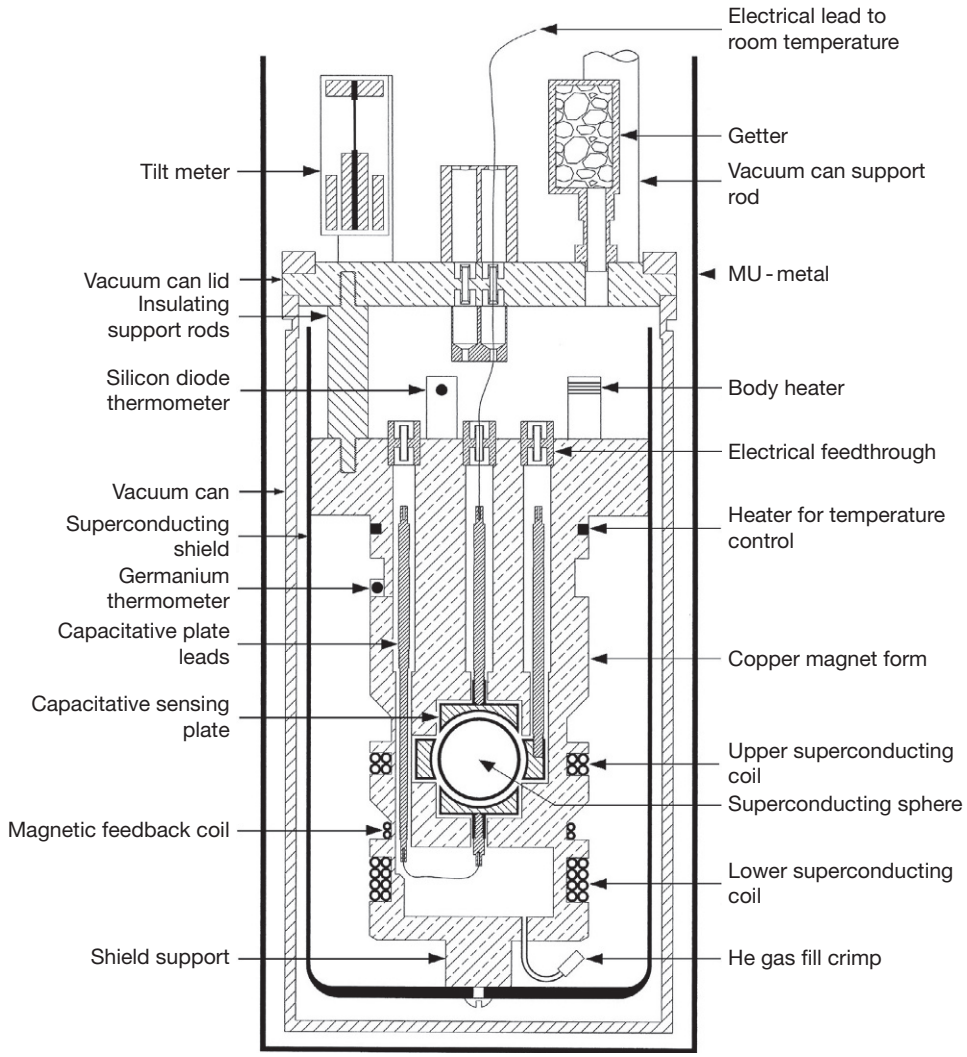


Figure 7 Schematic of SG sensor showing arrangement of the sphere, coils, vacuum can, and shielding.

are the levitated mass (sphere), the field coils, and the magnetic shield. The displacement transducer is located in a sealed cavity inside of the magnet form that supports the coils. It is formed by a capacitance bridge that surrounds the sphere and contains a partial pressure of helium gas. Two niobium wire coils (superconducting below a temperature of 9.2 K) carry perfectly stable persistent currents that generate an extremely stable magnetic field. The stability depends on the zero resistance property of superconductors. After the currents are 'trapped,' no resistive (ohmic) loses are present to cause them to decay in time. The test mass is a small 2.54 cm diameter sphere, also made of niobium, that weighs about 5 g. The coils are axially aligned; one just below the center of the sphere and one displaced about 2.5 cm below the sphere. When current flows in the coils, secondary currents are induced on the surface of the sphere, which by the Faraday induction law precisely cancel magnetic flux from entering the sphere. As with the currents in the coils, the induced currents are perfectly stable in the absence of any ohmic losses. The levitation force is produced by the interaction between the magnetic field from the coils and the currents induced on the surface of the superconducting sphere. Figure 8 shows a schematic of the sphere, coils, capacitance bridge, and magnetic flux lines.

The use of two coils allows the operator to independently adjust both the levitation force and the magnetic gradient. The upward levitation force is mainly produced by the lower coil. Its current can be precisely adjusted to balance the time-average downward force of gravity on the sphere at the center of the displacement transducer. The upper coil is used to adjust the magnetic force gradient ('spring constant'), which can be made very weak. As a result, a very small change in gravity (acceleration) gives a large displacement of the test mass, generating an instrument of very high sensitivity.

Because the levitation force is magnetic, changes in the Earth's magnetic field could seriously degrade its stability. To insulate the sphere and levitation coils from such changes, a superconducting cylinder with a hemispheric closure on one end surrounds them and is attached to the bottom of the copper magnet form. Furthermore, when the magnet coils are turned on, persistent currents are induced in the inside surface of the shield, which prevents the levitation magnetic field from penetrating the shield. An additional μ-metal shield is placed

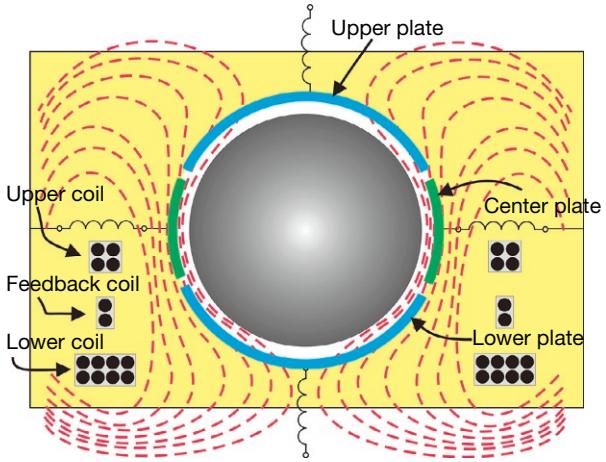


Figure 8 Schematic diagram of the Nb sphere, coils, plates, and general pattern of magnetic flux lines. Flux is excluded from the sphere and is confined externally by the Nb shield.

on the outside of the vacuum can. During the initialization process, this shield reduces the Earth's magnetic field by a factor of about 100 before the superconducting components cool through their transition temperature. This process minimizes any trapped flux in the sphere, coils, or shield, which could produce instability in the sensor.

3.04.1.2.2 Displacement transducer and feedback system

Relative motion between the ground (to which the coil assembly is attached) and the sphere – or any other perturbation of the gravity potential – moves the sphere from its equilibrium position. The position of the sphere is detected by using a phase-sensitive lock-in amplifier in conjunction with a capacitance bridge. Three capacitor plates surround the sphere with 1 mm clearance (Figure 7). The upper and lower plates are hemispheric caps that surround the upper and lower portions of the sphere. The center plate is a spherical ring around the equator of the sphere (Figure 8). A 10 kHz reference signal from the lock-in amplifier drives the primary of a carefully shielded transformer. The two balanced secondary windings of the transformer apply equal and opposite voltages to the upper and lower capacitor plates. The AC signal from the center ring plate is proportional to the displacement of the sphere from the center of the bridge. The sensor is operated in feedback to take advantage of the increased linear dynamic range and rapid response compared to open-loop operation.

The AC signal is amplified, demodulated, filtered, and applied to an integrator network. The DC output is connected to a precision resistor in series with a five-turn coil wound on the copper magnet form below the sphere. The resulting feedback force is proportional to the product of the feedback current and the current on the surface of the sphere. This force is given by $F = CI_F(I_{IC} + I_{IF})$, where I_F is the feedback current, I_{IC} is the current induced on the surface of the sphere by the levitation field, I_{IF} is the current induced on the surface of the sphere by the feedback field, and C is a constant. Since I_{IC} is proportional to g and I_{IF} is proportional to the amplitude of the tides, the maximum nonlinearity is $(I_{IF}/I_{IC})_{MAX} \sim 10^{-7}$. The sensor is therefore extremely linear. The gain (scale factor) of the sensor is determined by the geometry, the resistor size, the number of turns on the coils, and the mass of the sphere.

3.04.1.2.3 Temperature control

The sensor and superconducting shield are located inside a vacuum can surrounded by a liquid helium bath at about 4.2 K. In response to atmospheric pressure, the boiling point varies by about 1 mK mb^{-1} and during storms may change as much as 100 mK. Consequently, the sensor must be temperature-regulated (Goodkind, 1999). A germanium thermometer measures temperature and forms one arm of a Wheatstone bridge that has its null point preset to 4.5 K. The bridge output supplies feedback power of a few milliwatts, which is applied to a heater attached to the copper magnet form. Variations in control power almost perfectly follow the inverse of atmospheric pressure. With small bath temperature variations, high vacuum isolation, and high thermal conductivity of materials, it is relatively easy to regulate to a few microkelvin at cryogenic temperatures. As a result, the SG is almost completely isolated from environmental effects caused by changes in external temperature, humidity, and direct barometric pressure (but indirect atmospheric loading is always present). This is a major advantage over mechanical gravimeters that operate near room temperature.

3.04.1.2.4 Tilt compensation system

To measure the magnitude of the gravitational acceleration $\boldsymbol{g}$, the gravimeter must be aligned with a plumb line along $\boldsymbol{g}$. For most types of gravimeters, the test mass is constrained to move only along its axis of measurement. Therefore, when its axis is tilted with respect to the vertical plumb line, it measures only the component of $\boldsymbol{g}$ along its axis of measurement. The measured magnitude is $g\cos\theta$, where θ is the angle between the vertical and the axis of the instrument. For small angles, the apparent decrease in gravity produced by tilts is $\delta(\theta) = g\cos(\theta) - g \approx -g(\theta^2/2)$. The gravimeter is aligned along $\boldsymbol{g}$ by tilting it systematically along two orthogonal directions and setting it at the maximum value of $g(\theta = 0)$.

When an SG is tilted, the component of gravity along its axis of measurement decreases as $g\cos\theta$. The magnetic force gradient perpendicular to its axis of measurement is relatively weak, however, so the sphere moves off axis in response to the force component $g\sin\theta$. Because the magnetic levitation force supporting the sphere decreases off axis, the sphere position moves downward. This apparent increase in gravity has the same angular dependence as the equation in the preceding text, but its magnitude is about two times larger. As a result, the SG tilt dependence becomes $\delta(\theta)_{SG} \approx g(\theta^2/2)$ where the magnitude of $|g_{SG}| \approx |g|$. Therefore, the SG is aligned along $\boldsymbol{g}$ by tilting it systematically along two orthogonal directions and setting it at the minimum value of g (not at the maximum, as for other gravimeters). This phenomenon explains why disturbances due to local loading from nearby trains or automobiles will cause downward spikes on the SG signal. The horizontal accelerations move the sphere off axis where the magnetic support force is weaker. This effect was first observed in 1981 when the TT30 was installed in the cellar vault in the ROB. In contrast, trains did not affect operation of the LCR gravity meter operating in a nearby vault.

The SG is supplied with an automatic leveling system consisting of two tiltmeters mounted orthogonally on top of the gravimeter vacuum can and two thermally activated levelers that are placed under two of the dewar support points. After tilt-minimizing the SG sensor, the tiltmeters are 'aligned' to the same null by electronically setting their output voltages to zero. In feedback, the tiltmeters continuously adjust the power controlling the expansion of the levelers to keep alignment better than 1 μrad. This leveling precision is essential in gravity studies where apparent tilt-induced gravity changes must be kept <1 nGal. The tilt minimum adjustment is made on initial installation and checked every year or so by the operator. A study by Iwano and Fukuda (2004) on SG data from the Syowa station shows the clear advantage of the tilt compensation system in reducing the noise in gravity, especially in the tidal range.

Riccardi et al. (2009) analyzed a long data set (more than 11 years) of tilt signals observed on the Strasbourg SG. Most of the observed signals have a thermal origin dominated by a strong annual component of about 200 μrad but found also a small tidal component of about 0.05 μrad. They conclude that the automatic tilt feedback system device satisfactorily compensates for the tilt of the SG, whatever its origin, except for high-frequency (>1 mHz) perturbations.

3.04.1.2.5 Sphere and sphere resonance

The sphere is a hollow superconducting shell that is manufactured with a slight mass asymmetry so that it has a preferred orientation when levitated. Various manufacturing processes are discussed in Warburton and Brinton (1995). A small hole is drilled in the top of the sphere to allow the helium gas to enter and to prevent a differential pressure from developing when it is cooled to 4 K. Just as important, the volume of the sphere shell displaces 15 times less helium gas than the volume of the sphere. The hole thus reduces buoyancy forces that result from changes in the surrounding helium gas pressure.

When the gravimeter is tilted, the horizontal displacement of the sphere turns into an orbital motion (precession) with an associated vertical component in the feedback output. This mode appears as a sphere 'resonance' that has a period of 60–120 s depending on the particular instrument (Imanishi, 2005). In the absence of trapped magnetic fields and helium gas in the chamber, the Q of this mode is several thousands, so it is always excited, rendering the instrument not usable. Slow damping of the mode is provided by adding helium gas to the chamber, but the resonance remains underdamped and is clearly visible in some of the instruments' data. By comparison, the vertical resonance of the sphere is heavily damped with a period close to 1 s. Further technical details on the instrument design can be found in Goodkind (1991, 1999).

3.04.1.3 Instrument Performance

3.04.1.3.1 Instrument drift

One SG design goal was to produce an instrument that is stable to a few microgal per year. Meeting this goal with a commercial instrument took approximately a decade. Drift in the early model TT70s could be more than 100 μGal year^{-1} during the first year of operation but decreased steadily with time to <10–20 μGal year^{-1} after 5 years of operation. Drifts were modeled as a sum of one or two exponentials, a low-degree polynomial, or a combination of both (Boy et al., 2000; Seama et al., 1993), but uncertainty in the exact functional form decreased the precision with which long-period tides and polar motion could be determined from the data.

A dramatic decrease in drift was made in the early 1990s owing to a variety of design changes described in Warburton and Brinton (1995). At present, drifts are characterized by a small initial exponential followed by a small linear term. The exponential component decays in 4–6 months, after which it is negligible. For example, the initial drift for C023 manufactured in 1995 is $d(t) = -16.0\exp(-t/31)$ μGal, where t is in days (Schwahn et al., 2000). Long-term linear drift rates (including real secular changes) reported for the nine SGs operating in Europe vary from 1.6 to 4.9 μGal year^{-1} (Crossley et al., 2004). An 8-year comparison of an FG5 AG next to SG C023 in Membach confirmed that, after removal of the linear term (4.2 μGal year^{-1}), the SG has an identical spectra to the AG for frequencies <1 cpd (cycles per (solar) day) (Van Camp et al., 2005). This confirms that the SG drift is restricted to DC (very low frequency) and that it provides a continuous low-noise record of all gravity variations.

The only reliable method to determine instrumental drift is to compare the SG with collocated measurements made with an AG at regular intervals. The SG provides a complete time history of gravity at the site, which is invaluable for correlating with other geophysical variables, whereas the AG provides an absolute reference from which the drift and secular changes can be inferred. Precise drift measurement is complicated by real gravity variations due to hydrology, crustal uplift, annuals terms, seasonal terms, or signals of unknown origin (e.g., Zerbini et al., 2002).

3.04.1.3.2 Calibration stability

The question of calibration will be covered in more detail later in this report (see **Section 3.04.2.4.2**), but it is worth noting here that the recalibration of single-sensor instruments using AGs over time periods of 5 years or more has rarely shown differences that exceed the calibration error (between 0.1% and 0.01%). For dual-sphere sensors, the constancy of calibration is verified to 0.001% by least squares fitting the time series of the upper sphere to the lower sphere (Kroner et al., 2005). Similar experiments involving side by side comparisons of single- or dual-sphere instruments (Meurers, 2012; Wziontek et al., 2012) confirm constancy to 0.02%.

3.04.1.3.3 Instrumental noise and precision

Most of the GGP data acquisition systems record and digitize the full output of the SG electronics board. A voltmeter with 7.5 digits of resolution has the equivalent of 22 bits. When applied to a signal that includes the full range of tides (300 μGal), this translates into the smallest significant change of about 0.1 nGal. This is, in effect, the quantization noise of the SG and the data acquisition system, but not necessarily its precision. There is no reference gravity more accurate than the SG itself, so the precision of the SG can only be obtained by inference. In the frequency domain for studies of tides or normal modes, it is common for the SG to measure small periodic tidal signals and long-period seismic signals with a

sensitivity of 1 nGal and better. Therefore, 1 nGal is generally referred to as the nominal precision, or sensitivity, of the SG.

The determination of the instrumental noise of the SG is complicated by the fact that sources of the Earth noise (signals from the atmosphere, oceans, and hydrology) are generally much larger than instrumental noise for frequencies ranging from 3×10^{-8} Hz (1 cycle per year) to 1 Hz. The SG noise is higher than Earth noise only in the small subseismic band between 1 and 20 mHz, whereas the noise level is typically from 1 to 3 nm s^{-2} $Hz^{-1/2}$ (0.1–0.3 μGal $Hz^{-1/2}$) for most SG stations (Rosat et al., 2004). This is more than two orders of magnitude lower than the noise level of the AG as reported by Crossley et al. (2001) and Van Camp et al. (2005).

A comparative analysis on colocated records in Strasbourg (France) of SG-C026 and a new-generation spring gravimeter, the Micro-g LaCoste, Inc. gPhone-054, confirms that the SG has better performances on the whole analyzed spectral band, ranging from the body tides to the long-period seismic band (Riccardi et al., 2011). The instrumental drift observed on the spring meter during this study still remains a critical point compromising the study of long-term gravity changes.

It is common practice to filter and decimate data to 1 min samples to look at small temporal gravity variations. Assuming white noise, the aforementioned noise level translates into a precision of 0.01–0.03 μGal. This level is consistent with common experience. For example, Meurers (2000) easily observed gravity signals of magnitude 0.1–0.3 μGal over 10–30 min intervals, and Imanishi et al. (2004) identified reliable coseismic offsets of a similar magnitude.

The best way to determine real-world accuracy in the temporal domain is to compare different instruments operating side by side. Two comparisons have been done – one in Miami (Harnisch and Harnisch, 1995; Klopping et al., 1995; Richter, 1990) and one in Boulder (Harnisch et al., 1998) – concluding that different pairs of instruments agree to 0.1 μGal or better. For a dual-sphere instrument, the data streams from the two spheres are largely independent, and the difference signals between two sensors of the four dual-sphere SGs are typically 0.1 μGal over record lengths of years. Wziontek et al. (2012) determined that sensor differences account for up to 0.03 μGal at periods <1 h and rise to about 0.08 μGal at a period of about 1 day. Additionally, the residual curves from the dual-sphere SG in Moxa agree within a few tenths of a μGal to the polar motion modeled using data provided by the International Earth Rotation and Reference Systems Service (IERS). These results suggest that 0.1 μGal is the time-domain accuracy of the SG for long periods.

The instrumental noise power, $P(\omega)$, for a single-axis accelerometer can be expressed as $P(\omega) = 4kTb\, m^{-2}$, where m is the mass, k is the Boltzmann constant, T is the temperature, and b is the damping constant proportional to velocity. Therefore, increasing the mass will significantly decrease the noise. To test this concept, the most recent dual-sphere OSG-056 was manufactured with a heavier mass, a 17.7 g sphere in the lower sensor LG1, while a typical mass, a 4.34 g sphere, was used in the upper sensor UG2. In September 2009, OSG-056 was installed in the Black Forest Observatory (BFO) in Germany, and it has been operating continuously since its installation.

Figure 9 shows a comparison of the noise for OSG-D056 compared to the STS-1 seismometer also operating at BFO. The green line labeled UG2–αP shows the residual noise for the 4.34 g sphere after the atmospheric pressure, P, was removed using an admittance α of 3 $nms^{-2}hPa^{-1}$. The red and orange lines are data from the 17.7 g sphere. The red line LG1 shows data before removing atmospheric pressure, in agreement with the STS-1 data below 1 mHz. The orange line labeled LG1–αP shows the improvement from removing the atmospheric pressure. This shows that the noise of the 17.7 g sphere below 1 mHz is well below that of the 4.3 g sphere, STS-1, and NLNM.

At present, GWR is continuing to experiment with the SG sensor to further decrease its noise. One promising approach is increasing the Q of the SG sensor, which is typically <0.1. Since $b = m\omega_0/Q$, where $\omega_0 = (K/m)^{1/2}$, raising the Q could significantly decrease the noise. In an experimental sensor, capacitance bridge components near the sphere were replaced with nonconductive materials. These reduced nearby sources of Eddy current damping and resulted in Qs that ranged from 1 to 30, producing a 10–20 dB further decrease in the noise.

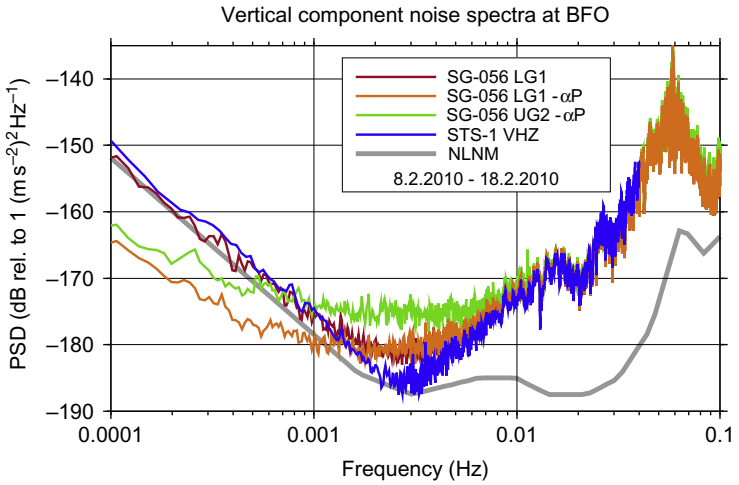


Figure 9 Comparison of noise power spectral densities of the dual-sphere OSG-D056 operating for ten quiet days at the Black Forest Observatory (BFO), Germany. Courtesy of R. Widmer-Schnidrig.

If successful, SGs could further contribute both to normal-mode seismology (Widmer-Schnidrig, 2003) and to the search for core modes and the Slichter triplet in the subseismic band (1–6 h) (Rosat et al., 2004). A table of the performance and noise characteristics (such as precision) of SGs compared with those of absolute and spring gravimeters can be found in the recent review by Crossley et al. (2013).

3.04.1.4 User Requirements

3.04.1.4.1 Operation and maintenance

The SG sensing unit contains only one moving mechanical part, the niobium sphere. It is therefore virtually free of any maintenance requirements, as verified by field installations of 15 years. SG support equipment, however, does need periodic maintenance to assure proper operation and can fail unexpectedly as the result of a lightning strike or other natural catastrophe. Many of the major gaps in SG data have been caused by power supply failure during major storms or failure of the data acquisition systems. Planning for failure of either electronics or refrigeration is necessary to minimize interruptions in long (decadal) gravity records. It is most important to keep the dewar at least partially filled with liquid helium, so the sensor and superconductors remain below 4.5 K. Upon complete helium loss, the sensor will start warming up to room temperature. Although no damage occurs to the sensor, it requires that the sphere be relevitated, which reactivates the initial drift discussed in **Section 3.04.1.3.1**. In practice, therefore, operators are very careful to make sure that the liquid helium volume is kept above a minimum level, so in the case of a power or cold head failure, there is enough time either to transfer more liquid helium or to fix the source of the failure. With the cold head off, the maximum hold time for a compact dewar is about 60 days. Prudently, most operators do not let the liquid He fall below one-third. Therefore, even under severe interruptions, such as the fire at Mt. Stromlo, Australia, in January 2003, the operator has at least 20 days to solve the resulting problems without warming the sensor up. It is also important to follow the manufacturers' and GWR's instructions for maintenance of the cold head, compressor, and water chiller to prevent equipment failure. Many operators keep a backup refrigeration system available for immediate replacement.

At most of the GGP stations, operators check weekly that the refrigeration system and data acquisition system are functioning properly and ensure general site integrity. Problems that develop either are observed in the support status variables that monitor operation of the support equipment (temperature control, tilt-leveling control, and the refrigeration system) or will cause an increase in the instrumental gravity noise. For example, refrigeration problems cause immediate increase in helium boil-off rate and warming of dewar neck thermometers. Ice buildup around the cold head that touches the inside of the dewar neck will cause an immediate increase in noise observed through the mode filter and on the gravity residual. Problems with the leveling system will be observed on the Tilt X and Y balance signals and as gravity noise on the mode filter and gravity residual.

The new GWR data acquisition system (DDAS) allows the operator to monitor about 30 status variables remotely. In addition, alarm levels can be set to automatically generate warnings and alert the operator by e-mail to initiate investigation and repair. After collection and analysis of 1 month's data, the operator can enter a calibration factor, tidal parameters, and barometric pressure admittance, and the DDAS will automatically generate a theoretical tide and display the gravity residual signal in real time. This allows visual examination of the gravity noise at the sub-μGal scale. Changes in noise level are immediately observable and with some experience can be identified as those of geophysical origin (atmosphere, ocean, or earthquakes) or those caused by possible equipment problems. In the latter case, GWR can consult online with the operator to analyze the problem and provide a rapid solution. Remote access should reduce the frequency of data gaps and ensure higher quality of overall long-term data.

3.04.1.4.2 Site location

The SG measures an extremely wide bandwidth of signals, from periods of seconds to years, and the origin of signal sources ranges from local to global. Site selection is paramount to determine the signals of most importance to the scientific objective. Thus, a site may be chosen to maximize gravity signals of interest while minimizing signals of less interest that would be considered 'noise.' For example, a site must be near a volcano if one wishes to measure signals from magma intrusion, and it is better to be close to an ocean to measure ocean loading or sea-level changes. In contrast, if the goal is to measure short-period signals such as seismic waves, normal modes, or tides, the site needs to be as quiet as possible (see the succeeding text).

Site selection may be restricted to the country providing the research funds and by the goals of the funding agency; however, some sites are cooperative efforts chosen to expand the geographic distribution of the GGP network. Examples of the latter include Ny-Ålesund, Norway (Sato et al., 2006a); Sutherland, South Africa (Neumeyer and Stobie, 2000); and Concepción, Chile (Wilmes et al., 2006). Neumeyer and Stobie (2000) discussed both geographic and practical criteria used for choosing the Sutherland site. The SG has been used in a wide variety of field situations, including harsh conditions such as a salt mine in Asse, Germany (Jentzsch et al., 1995), and on the edge of the Antarctic continent at Syowa Station (Sato et al., 1991). Most of the sites, however, are more instrument-friendly and have been chosen to exploit existing facilities where other geophysical instrumentation is already operating, sharing infrastructure (such as power, telephone, or satellite communications) and staff.

3.04.1.4.3 Site noise

The SG is an extremely low-noise instrument and requires a quiet site for optimum operation. A quiet site is one that is removed, by distances of at least several kilometers, from nearby anthropogenic environments, such as a city, town, railroad, or major highway. These are strong noise-generating environments that not only increase short-period disturbances but also introduce ground tilts and loadings that will be seen in the residual gravity. If possible, potential sites should be pretested for ambient Earth noise using an STS-1 VBB long-period seismometer. To detect low-noise signals, the site noise must approach the new low-noise model of Peterson (1993) and the

best IRIS Global Seismographic Network stations (Berger et al., 2004; Widmer-Schnidrig, 2003).

Note that a site may be isolated from anthropogenic influences but may still be part of a scientific research station containing large (and massive) instruments such as a VLBI (very long-baseline interferometry) antenna. These fiducial stations are of great interest from geodetic and geophysical points of view, due to the advantages of combining gravity and geodetic measurements. If, however, an SG is housed within a busy scientific building that is visited constantly or host to other machinery, the quality of the data will suffer. Ideally, the SG should be housed by itself some distance (100 m or more) from other instruments or heavy-traffic areas and should pass the seismometer test proposed in the preceding text.

3.04.1.4.4 Site stability

The monitoring of long-term crustal deformation using gravity and space techniques requires careful integration of GPS, AG, SG, environmental sensors, and ocean gauges (Richter et al., 2004; Zerbini et al., 2001, 2002). All piers must be constructed carefully, since differences in pier construction and the separation of piers (noncollocation) may cause subtle and spurious signals of nongeophysical origin. The deformation characteristics of the ground immediately below and around the instrument are obviously important. Ideally, if an instrument can be placed on a concrete pad that is anchored directly to nearby nonfractured bedrock, deformations at the SG will reflect those over a much larger surrounding area. SGs are frequently located in an underground setting together with other instrumentation (seismometers and tiltmeters), normally on bedrock. There is no problem in siting an SG within such a complex, provided that an environment is created around the instrument to protect it and the associated electronics from excessive humidity and temperature changes. Since the cold head and compressor may be too noisy to be placed near other instruments, they must be sited with some care.

For aboveground situations, it may not be possible to find bedrock, and the gravimeter will have to sit on a concrete pad secured to unconsolidated material. The type of material (clay, gravel, sand, etc.) will play a role in the interpretation of meteorologic and other seasonal effects. For example, it has been found that porous material can compress and deform under loading and thus generate an unwanted signal. Whatever the height of the SG with respect to the local ground level, an important factor effect is the soil moisture content of the local and regional areas from 1 to 100 m around the gravimeter. Soil moisture resides largely in a layer no more than 1 or 2 m thick, but the effect on an SG can be significant. Installation of groundwater sensors, soil moisture probes, rain fall gauges, and other meteorologic sensors is required for developing advanced hydrologic models at the sub-μGal level.

The requirements of location, site noise, and site preparation need to be assessed very carefully by potential new users. Considerable experience lies both with the manufacturer (GWR) and with many experienced SG owners, who have maintained their stations for a decade or longer. Although there is no central funding source for establishing new SG sites, potential new users are encouraged to contact existing SG groups for advice, particularly through the GGP website, through GGP workshops, and by visiting operating GGP sites.

3.04.2 SG Data Analysis

As with all modern geodetic techniques, data from SGs require specialized processing before they can be used to their fullest advantage. There is a large amount of GGP data available from GGP-ISDC (Information System and Data Center) online (http://ggp.gfz-potsdam.de/), hosted by ICET (International Center for Earth Tides), but they may be of limited use to scientists unfamiliar with this kind of data. The purpose of this section is to review the common procedures in analyzing the data for different end uses. A summary of the GGP project is given in Crossley et al. (1999) and Crossley and Hinderer (2009); station names and details can be found in Crossley (2004).

3.04.2.1 Preprocessing

3.04.2.1.1 Second data sampled from an SG

SG data acquisition systems usually record two types of signal: the gravity feedback signal at high accuracy (0.1 nGal) and at precise times and many auxiliary channels such as environmental data (e.g., atmospheric pressure and room temperature) and instrument parameters (e.g., tilts). The gravity is typically recorded at 1, 2, or 5 s intervals, whereas the auxiliary signals are often sampled at lower rates, for example, 1 min. We note that typically users take advantage of the GGP antialiasing filters that permit 1 or 2 s sampling, but other (unfiltered) signals can be accessed from the SG electronics that permit sampling at a higher 10 Hz rate, discussed in Van Camp et al. (2008).

Examples of raw SG data are shown in **Figure 10**. The upper panel shows data from the Cantley and Boulder instruments at 1 and 5 s sampling, respectively. There is a noticeable difference in high-frequency noise between the recordings due to the different sampling rates and antialiasing filters. Most of the noise is microseismic, that is, the incessant propagation of surface waves generated by wind and ocean turbulence, between 1 and 10 s, in the period range where most SG data are sampled. The lower panel shows data from the Strasbourg instrument for the day after the Sumatra–Andaman earthquake (26 December 2004). The earthquake clearly dominates the tides, but SGs are not gain-ranging instruments, so the surface waves from large events are frequently clipped.

The second data are usually not sent to ICET, although there is a mechanism for receiving them. Initially, the issue was one of file size, but also, the high-rate data are not generally interesting from a geodetic point of view. Some of the raw data are transferred to the IRIS (Incorporated Research Institutions for Seismology) database where it is of interest in the study of seismic normal modes of the Earth (Widmer-Schnidrig, 2003).

3.04.2.1.2 Minute data from the GGP database (ICET)

Most of the data available at ICET/GFZ are minute data sent by each station on a regular basis (one file per month). The requirements of GGP are that operators must apply a digital decimation filter to their raw data and resample it at 1 min

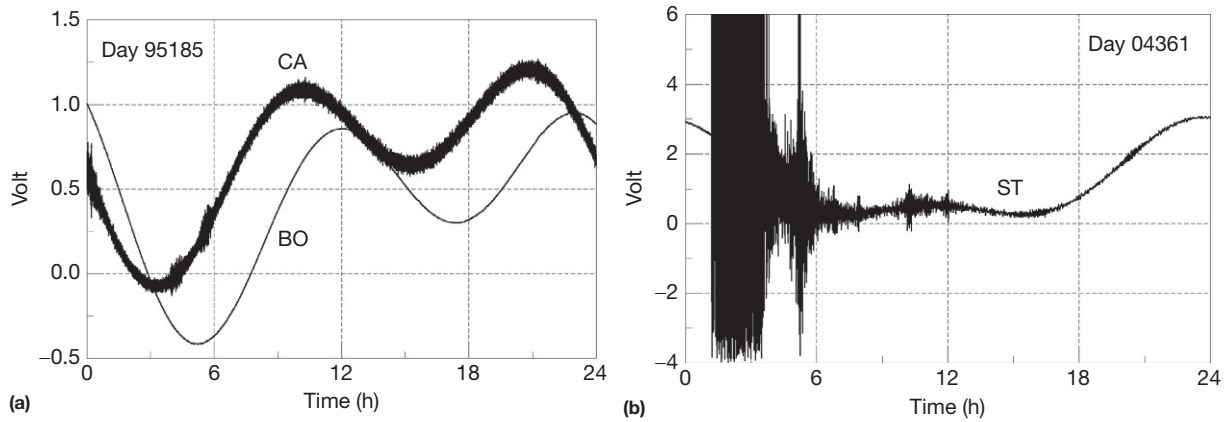


Figure 10 Examples of 1 day of raw SG data (a) from CA (Cantley, Canada – 1 s sampling) and BO (Boulder, United States – 5 s sampling) stations and (b) from ST (Strasbourg, France – 2 s sampling) station that includes the $M_w = 9.3$ Sumatra–Andaman earthquake.

Table 1 Sample GGP 1 min data file

```
Filename              : ST041200.GGP
Station               : Strasbourg, France
Instrument            : GWR C026
Phase Lag (sec)       :   17.1800       0.0100    measured
N Latitude (deg)      :   48.6217       0.0001    measured
E Longitude (deg)     :    7.6838       0.0001    measured
Height (m)            :  180.0000       1.0000    estimated
Gravity Cal (nm.s-2/V): -792.0000       1.0000    measured
Pressure Cal (hPa/V)  :   22.2222       0.1000    estimated
Author                : jhinderer@eost.u-strasbg.fr
yyyymmdd hhmmss gravity(V) pressure(V)
C*************************************************************
77777777
20041201  000000  2.448085  4.914420
20041201  000100  2.452300  4.912670
20041201  000200  2.456466  4.910337
20041201  000300  2.460378  4.908314
20041201  000400  2.464132  4.906599
```

using a zero-phase digital filter. The local air pressure is also decimated to 1 min and the two series are combined in a simple ASCII file; each sample is date-stamped and time-stamped and the file has appropriate header information. A sample of the data from the Strasbourg station is shown in Table 1 in the PRETERNA format that is part of the ETERNA Earth tide analysis program (Wenzel, 1996b). These data are uncorrected; that is, data disturbances such as spikes, offsets, and other problems still remain, inevitably smoothed by the decimation filter.

A display of minute data from ICET is shown in Figure 11; this is from station Boulder for the month of September 1997. The month is chosen for illustration purposes because it has data gaps, spikes, and (not visible at this resolution) signals that seem to be offsets. Generally, the data at ICET vary from bad months like this to completely trouble-free data; a typical file may contain one or two problems to be fixed. Each monthly file is about 1.5 MB uncompressed.

Two different philosophies have been used to solve problems in the data: either leave gaps (i.e., simply ignore bad segments) or remove disturbances and offsets and fill the gaps with a synthetic signal from a model. The former requires that all processing steps have to maintain the integrity of the gaps, and this reduces the flexibility of using most time series analysis algorithms in their standard form. The latter requires further choices about what level of disturbances to treat and what to leave alone, but it is the preferred approach within the gravity community. The main reason is the convenience of dealing with continuously sampled data rather than data in sequences of blocks and gaps. Computer algorithms are also easier to implement for continuous data than data organized into blocks, but this is not the only reason to repair bad data.

Some geophysicists would argue, with justification, that one should not 'invent' data for the convenience of processing. Pagiatakis (2000) proposed the treatment of SG data using a least-squares inversion for all possible constituents at once, as is commonly done for other complex data sets, for example, VLBI. In passing over 'bad' data, the amplitude of offsets is included as a variable in the inversion, so they are removed simultaneously with tides, pressure, drift, and other modeled effects. This

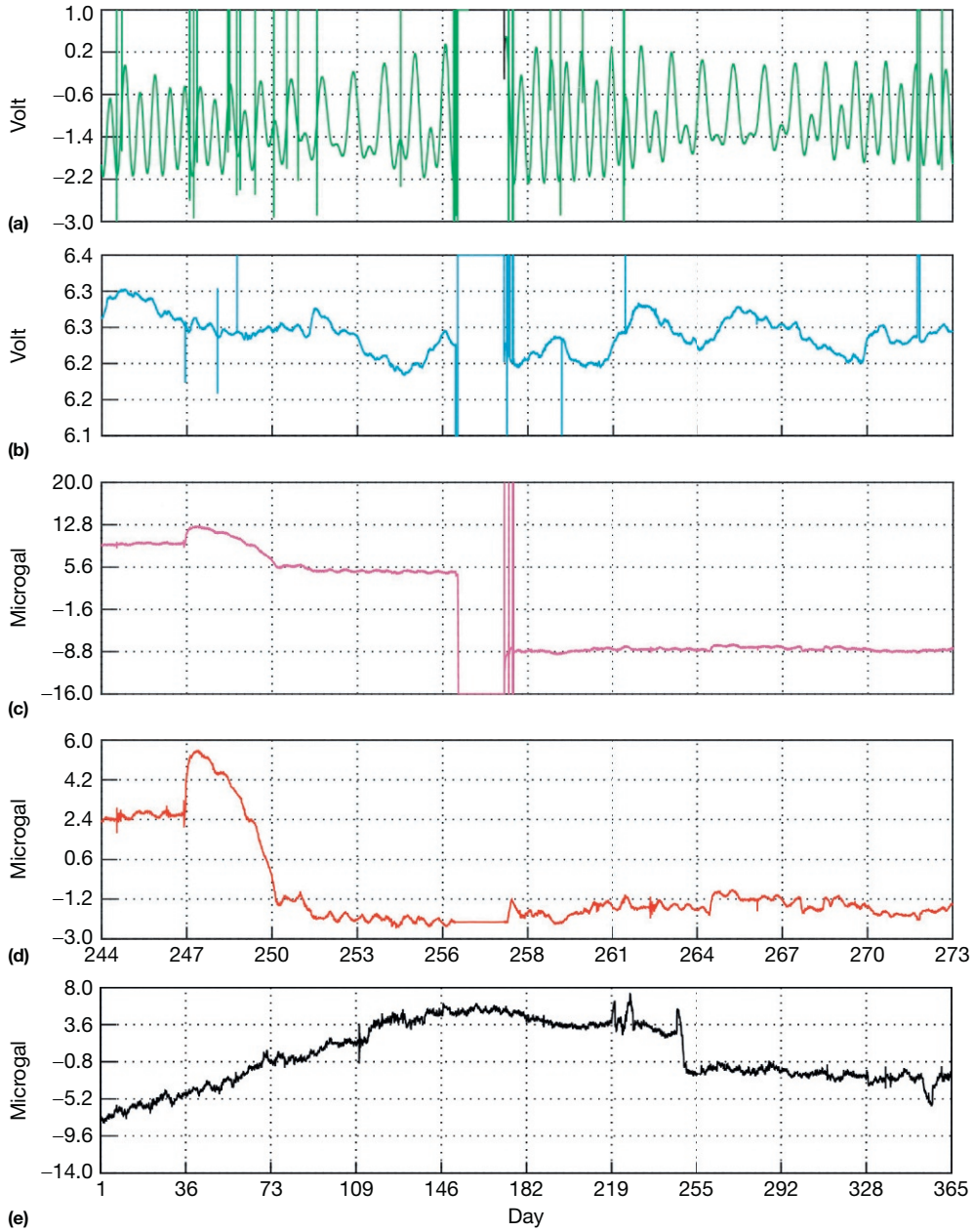


Figure 11 Preprocessing of 1 min data for BO station (September 1997). (a) Original gravity and (b) original pressure, (c) after small gaps (<10 s), are filled, and removal of tides and nominal pressure, (d) after a large gap with offset, is corrected, and (e) the residual gravity for the whole of 1997 after gaps and offsets is removed.

obviously requires a large computational effort, particularly when using 1 s data (as opposed to 1 min data) and when the data span many years or even decades. The inversion also requires interpretation of the resulting parameter covariance matrix to see what trade-offs exist in the model. This approach does not give an intuitive feel for the effects of various processing steps, and there are many different corrections to be included. For the rest of the chapter, we will assume that the goal is to produce and analyze a uniformly sampled residual gravity series.

A simple model will suffice for most of the following discussion:

$$\begin{aligned} g(\text{observed}) = {} & g(\text{disturbances}) && (\text{of instrument and station origin}) \\ & + g(\text{tides}) && (\text{solid Earth, ocean}) \\ & + g(\text{nontidal loading}) && (\text{atmosphere, ocean currents}) \\ & + g(\text{polar}) && (\text{annual and Chandler polar motion}) \\ & + g(\text{drift}) && (\text{instrument drift function}) \\ & + g(\text{hydro}) && (\text{rainfall, soil moisture, groundwater}) \\ & + g(\text{residual}) && (\text{other signals, deformation, tectonics,} \\ & && \text{slow earthquakes, etc.}) \end{aligned} \quad [1]$$

The purpose of data processing and analysis is to remove the effects that are not the subject of investigation and model the effects that remain. We note at this point that the instrument amplitude calibration factors are required in the next step

in order to convert the observed (g, p) in volts to their equivalent gravity and pressure values:

$$g\,(\mu\text{Gal}) = g\,(\text{volt}) \times \text{GCAL}$$
$$p(\text{hPa}) = p(\text{volt}) \times \text{PCAL}$$

Calibration issues are discussed in detail later in this chapter.

3.04.2.1.3 Remove–restore technique

The most widely used approach is to first make nominal corrections for some of the largest influences such as tides and pressure (remove), then fix the problems in the residual signal, and finally put back the removed signals (restore). The data can then be used for further processing, and analysis, as required. Before any corrections are made to gravity, the first step is therefore to deal with problems in the atmospheric pressure, because this will be used in the remove–restore phase. Figure 11(a) and 11(b) shows the gravity and pressure for 1 month; clearly, the pressure has disturbances and gaps at the same time as in the gravity channel, indicating problems in the data acquisition system.

With pressure data, it is difficult to generate data within gaps, short of running computationally intensive weather forecasting codes or interpolating data from meteorologic stations surrounding the gravimeter. Meteorologic services give hourly pressure data that are available by request or through the Internet. For simplicity, a linear extrapolation between the beginning and end points of the gap, even for gaps extending for days or more, may be sufficient. Pressure variations have only a small seasonal signal and no significant daily variation, so this interpolation is at least reasonable, if not ideal. For longer gaps of several days or more, an auxiliary source of data as mentioned in the preceding text should be considered, especially where the pressure channel has failed but the gravity data are still good.

Pressure sensors are usually factory-calibrated, but this calibration factor needs to be monitored by the user. From time to time, the sensors need recalibration for drift or other problems, or even replacement, and this can introduce a dilemma in the processing. Having detected a problem, should one allow for sensor drift over a previous period of time or simply assume an offset? No one solution is ideal in all cases. We assume that the pressure record has been made continuous through gaps and been cleaned of spikes and other disturbances. Depending on the circumstances, this might have required ancillary data or perhaps linear interpolation.

The gravity signal is not so easily dealt with. The data in Figure 11(a) are dominated by the tidal signal that varies usually between 100 and 300 μGals in amplitude, depending mainly on station latitude and the phase of the lunisolar cycle. For very short gaps or spikes (10 s or less), linear interpolation will work without much problem even without removing the tides. This is especially effective if applied to the original raw data (at second sampling) because spikes or gaps are not then smeared into the record by the second-to-minute decimation filter. For longer gaps, we need a version of the remove–restore philosophy.

3.04.2.1.4 Treatment of gaps in gravity data

To repair gravity data, we normally construct a preliminary model based on the tides and nominal atmospheric pressure. For longer gaps, it may be appropriate to add instrument drift and perhaps polar motion. The tides are usually modeled by summing wave groups with specific gravimetric factors (δ, κ) determined in some prior tidal analysis at the station (see the succeeding text). The preliminary model is then subtracted from the observed data to reveal hidden problems that need to be fixed. We call the residual gravity g (temp)

$$g(\text{temp}) = g(\text{observed}) - g(\text{removed}) \quad [2]$$

where

$$g(\text{removed}) = g(\text{tides}) + g(\text{nominal pressure}) + g(\text{drift}) + g(\text{polar motion}) \quad [3]$$

The instrument calibration factor already enters eqn [2]. Note that a linear function is normally used to fill in a residual gap, so we are interested only in modeling signals that are not linear during a gap. Polar motion is very well defined by IERS data, but at a level of 5 μGal year^{-1} (or 0.01 μGal day^{-1}), it is not important for gaps of less than a week or so. Instrument drift over the time period of most gaps can usually be taken as a simple linear function of time. Other components of the signal such as hydrology or nontidal ocean loading are, however, too complex to be included as part of the removed function, even though they can have an effect of several microgal over time spans of days.

Figure 11(c) shows g (temp) for the case when g (removed) consists of a local tide model + a nominal pressure correction. All the short spikes were previously replaced by a simple linear interpolation on the original data, as discussed in the preceding text. Because the pressure is included in the g (removed) model, it is corrected before the gravity in order to avoid introducing spurious signals into the gravity. The gravity series still has a major problem between days 256 and 258 but otherwise looks 'better.'

At ICET/GFZ, the basic files for all stations have uncorrected minute data of the form of Figure 11(a) and 11(b), that is, raw data decimated to 1 min but with no data repair prior to decimation. ICET produces corrected minute data in which data repair has been done on this already-decimated minute data. Some users also send data that have been repaired at the second sampling rate prior to decimation.

3.04.2.1.5 Disturbances and offsets

To make further progress, we now work with the residual series g (temp) as in Figure 11(c). The major disturbance at day 256 contains an offset, and there is a suspicious signal at day 247 that requires consideration. Repairing (fixing) the offsets in Figure 11(c) is one of the most critical processing steps, requiring choices as to which problems to leave alone and which to fix. We consider an offset (or tare) to be any apparent change in the instrument base level of nongeophysical origin. This could be a small offset (0.5 μGal) that takes place within one time step (e.g., 1 s) due to an instrumental or electronic event. Alternatively, an offset may occur during a helium refilling (maybe 10–20 μGal) or a major electrical strike that causes level changes of 100 μGal or more.

Here, we choose to fill the gap between days 256 and 258 with a straight line and to absorb the apparent level change into an offset that is 13.816 μGal. We have no means of knowing the actual offset during this time (the gap), so we simply

join the good data before and after with a straight line and compute the offset needed to minimize the residual signal. This is a large offset that is clearly associated with something that happened to the instrument, involving also the pressure; the instrument base level was changed by a nongeophysical event. Unfortunately, the data log is unavailable for this time period at Boulder, but at most stations, there should be a written explanation of such events (stored in the GGP database as 'log' files). Some disturbances will not cause an obvious offset in the data, but clear offsets should be removed.

The 'fixed' residual is shown in **Figure 11(d)** that interpolates the gap, but there is still an anomaly at day 247 with a jump of 2 μGal followed by a gradual decrease of 4 μGal. We know that rapid changes of gravity can be caused by physical processes such as heavy rainfall (Crossley and Xu, 1998; Klopping et al., 1995), but there is no reason at this point to adjust the data, even though the disturbance causes an apparent offset. Even though a log file helps to indicate a nongeophysical cause for an offset, it usually cannot help to decide what the size of a particular offset may be.

After going through the whole year 1997, we find a total of 29 problems that need to be fixed, three of which were readily apparent offsets. The residual curve for the whole year is shown in **Figure 11(e)**. Day 247 still shows up as a possible nongeophysical anomaly, as there is no apparent subsequent recovery from the level jump. Without the 'offset,' the data would show a smooth annual variation that arises from polar motion. Although it might be tempting to treat more aggressively the level change as an offset, doing so will affect all the subsequent data, and it is clearly risky to do so without good reason.

3.04.2.1.6 Automatic procedures

Ideally, each day's raw data record should be scanned visually for potential disturbances. Where this is not possible, simple numerical algorithms can scan the data for the rapid detection of anomalies. One method uses the slew rate (Crossley et al., 1993) that computes the simple forward data derivative and flags data points that exceed a certain threshold. We indicate from typical records the maximum slew rates in **Table 2**, although these may be exceeded in some situations. With the solid tides in the gravity signal, slew rates can reach 1 μGal min^{-1}, and this rate is important when trying to determine the phase calibration of the instrument, as discussed later. The next largest slew is from the atmosphere, and this is an order of magnitude larger than the slews from groundwater, rainfall, and soil moisture. Disturbances and offsets can have much larger slew rates, but sometimes, they are small slews that cannot be distinguished from the real signals in the preceding text. Most users would use slew rate as a diagnostic but not as a corrective procedure. Another approach is to compare each data point with a predicted value based on the accumulated statistics of previous data.

Table 2 Maximum slew rates for some geophysical signals

	(μGal min^{-1})
Solid earth tides	0.95
Atmospheric pressure	0.20
Groundwater variations	0.02

Two widely used software packages, Tsoft (Van Camp and Vauterin, 2005) and PREGRED (part of the ETERNA package, Wenzel, 1996b), deal especially with the preprocessing of gravity data, provide some automation of data analysis, and incorporate many other processing options. Tsoft is probably the most widely used package for GGP data.

To give further insight into the effect of processing choices, we show some examples from Hinderer et al. (2002a). In each case, the same 6 months of data (March–December 1997) was given to different scientists to repair, and the results were compared. **Figure 12(a)** shows how four different operators approached the problem of a moderate-sized earthquake that was removed. The results ranged from removal of just the large surface waves (SR) to removal of most of the disturbed record (JP). For a transient disturbance without offset, four different choices were made about which parts of the signal to repair (**Figure 12(b)**); depending on the length of data replaced, differences can be seen between the initial and final levels of the each of the processed series. Finally, a highly disturbed portion of the record due to a helium refill (**Figure 12(c)**) was an opportunity for two operators (JPB and DC2) to remove an apparent offset, and two others chose to leave it in. Some of this processing was done using Tsoft and some using other algorithms and procedures.

3.04.2.1.7 Processing for different purposes

For several end purposes, the long-term trends in the data are unimportant. This is the case for tidal processing, that is, for finding gravimetric tidal factors, ocean loading, and related signals such as the resonance effect of the free core nutation (FCN). Other possibilities are signals in the 1–24 h range (e.g., the Slichter triplet) and of course high-frequency seismic modes. In these cases, a more aggressive treatment of disturbances and offsets might be beneficial. **Figure 13(a)** shows the effect of the different processing options from **Figure 12** in the frequency domain, that is, the PSD (power spectral density) function. There are five different options ranging from a very minimal treatment (DC1 – not shown in **Figure 12**) to a comprehensive removal of problems (JPB). The latter has effectively reduced the noise floor in the long-period seismic band (up to about 5×10^{-5} Hz or 5.5 h) by more than an order of magnitude. This is mainly due to the removal of offsets and disturbances that contain high frequencies.

When the goal is to preserve the long-term evolution of the gravity residual rather than search for short-period signals, offsets must be examined carefully. If offsets occurred randomly in time and with random amplitudes, their accumulated impact is similar to a brown noise or random walk type of process. The mean value drifts from the true gravity, and this would be detectable only by comparison with absolute gravity measurements. Large offsets (5–10 μGal or more) need to be corrected, especially when they are logged as due to specific causes. Sometimes, smaller offsets happen without apparent cause, but if not corrected, a bias can easily accumulate. Even with appropriate software, this can be a time-consuming process, already discussed in the preceding text.

We show in **Figure 13(b)** the effect of different options in the time domain, over the 6-month test period referred to earlier. All the traces coincide at the beginning but

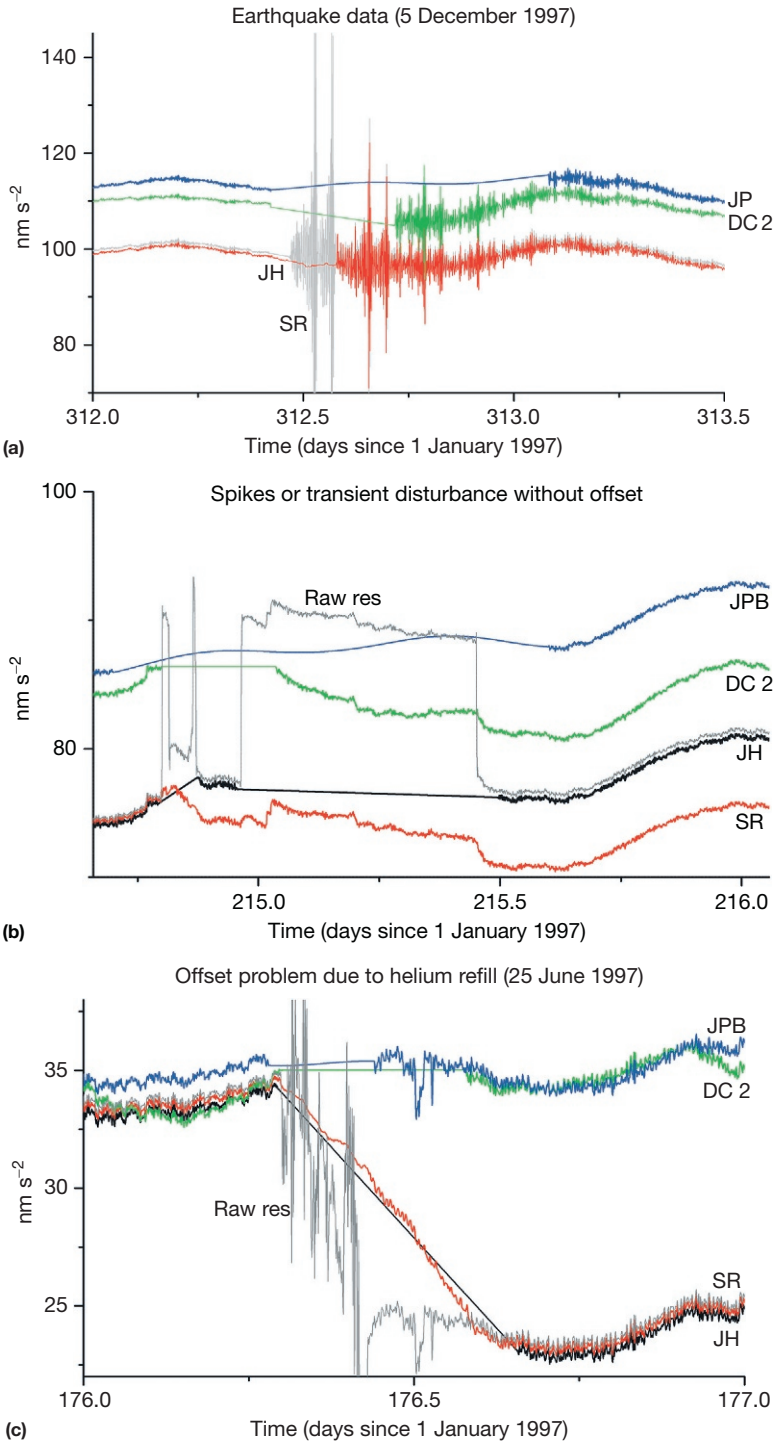


Figure 12 Examples of the data corrections done independently by different operators (identified by initials) for the removal/interpolation of (a) an earthquake (b) disturbances without offsets, and (c) an offset caused by a helium refill. Reproduced from Hinderer J, Amalvict M, Crossley D, Rivera L, Leveque J-J, and Luck B (2002). Tides, earthquakes and ground noise as seen by an absolute gravimeter and its superspring; A comparison with a broadband seismometer and a superconducting gravimeter. *Metrologia* 39: 495–501; Hinderer J, Rosat S, Crossley D, Amalvict M, Boy J-P, and Gegout P (2002) Influence of different processing methods on the retrieval of gravity signals from GGP data. *Bulletin d'Informations des Marees Terrestres* 123: 9278–9301.

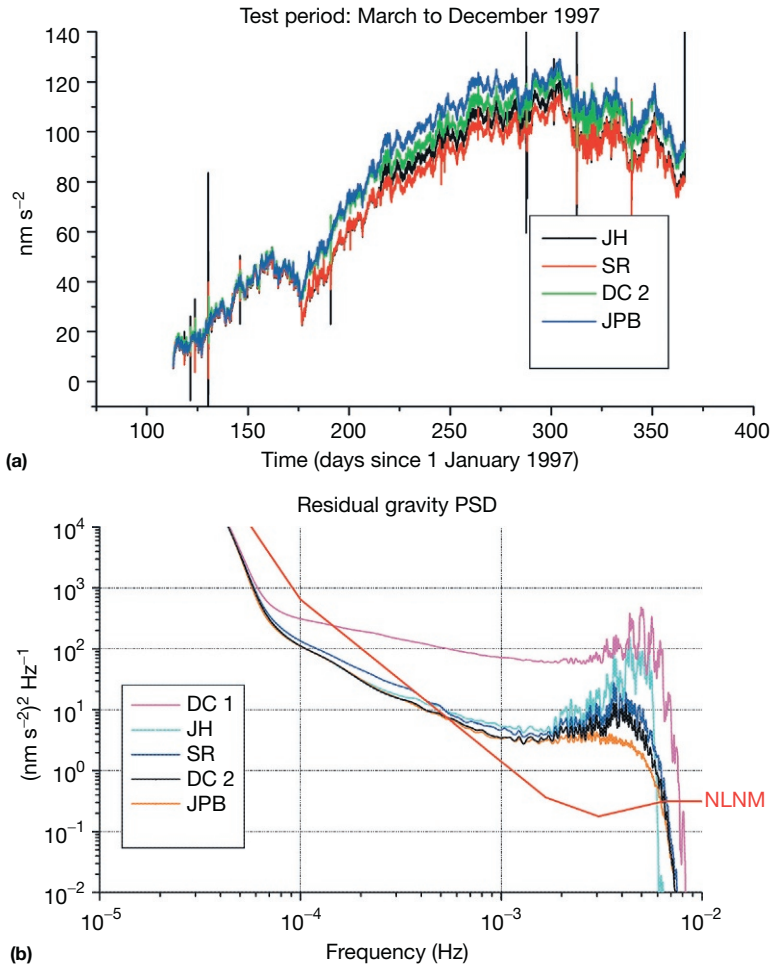


Figure 13 The cumulative effect of data corrections for a test period of 6 months of data from different treatments (a) in the time domain, showing a 1.5 μGal offset between treatments, and (b) in the frequency domain – NLNM is the new low-noise model of Peterson (1993).

diverge by up to 1.5 μGal after 6 months, the largest part of this difference being due to the different treatments of the disturbance in Figure 12(c). As indicated in the preceding text, the use of the AG measurements can be very useful in identifying offsets at the level of a few μGal. Note that in Figure 13(b), instrument drift plays no role in the processing differences.

Long gaps in a record make determination of offsets difficult if not impossible. There are two saving strategies – to rely on the collocation of AG measurement to anchor the SG record at the ends of a long gap and possibly to use a fit to the polar motion when there is a suspected large offset within the down time of the gravimeter. Long gaps in the record are generally due to serious problems in the instrument or data acquisition system, so an offset is certainly possible. Unfortunately, other instruments (e.g., spring gravimeters), even when available, have too much drift to provide a good interpolation signal to fill in an SG record for more than a few hours.

3.04.2.1.8 Restoring the signal

When all problems have been removed from g (temp), it is easy to recover the observed gravity from eqn [2], so that

$$g(\text{fixed}) = g(\text{temp}) + g(\text{removed}) \quad [4]$$

is now the observed gravity without the problem segments. Referring to eqn [1], we note that the fixed data have a signal in the gaps that does not contain the more complex effects such as hydrology. In all other respects, however, the data in the good portions are identical to the observations and the fixed gravity can now be analyzed for the components in eqn [1]. By subtracting g (fixed) from g (observed), it is always possible to see exactly what has been removed during this stage of the processing.

3.04.2.2 Solid Earth and Ocean Tides

3.04.2.2.1 Tide-generating potential

The basis for computing a theoretical gravity tide is the tide-generating potential (TGP). Using the orbital and rotational data for the Earth as forced by the Sun and Moon, this potential was first given as a catalog by Doodson (1921) who included only terms of degree 1–3 (i.e., 24, 12, and 8 h periods). The tides occur in the rotating frame of the Earth's mantle, but they also lead to forced nutations of the Earth in the space-based

(nonrotating) coordinate system (e.g., Melchior and Georis, 1968). There is one tidal wave associated with each component of the Earth's nutation. The most extensive tidal developments now include the perturbation effects of all the major planets and terms up to degree 6 for the Moon (period 4 h) and terms allowing for the nonspherical shape of the major bodies. The most recent references on the tidal potential for gravity work are Hartmann and Wenzel (1995a,b), Roosbeek (1996), Wenzel (1996a), and Kudryavtsev (2004).

There are two approaches to cataloging the TGP. First, the ephemeris is a catalog of apparent positions of the bodies in the solar system as seen from a position and time on the Earth; each body is defined by a longitude, latitude, right ascension, and obliquity. In the first approach, an ephemeris tide program, such as GTIDE (e.g., Merriam, 1992a), generates a time series from the TGP on the surface of a rigid Earth. The accuracy of this TGP is directly related to both the accuracy of the ephemeris itself and the analytic theory used to describe the motions of the planetary bodies. Merriam estimated for GTIDE an rms time-domain accuracy of 0.25 nGal, with a maximum error of 0.8 nGal.

Wenzel (1996a) compared three tidal ephemeris programs: GTIDE as updated in Merriam (1993) and two JPL programs DE200T and DE403T that he used to generate six benchmark gravity tide series. These series were gravity at 1 h spacing for two example epochs (1987–93; 2017–23) at the BFO in Germany (chosen because of the availability of gravimeter data at BFO). The internal time-domain accuracy of the benchmark series derived from the JPL ephemerides is better than 0.1 nGal, and in the frequency domain, the rms errors are about 100 times smaller.

The second approach, which is a more widely used method for the TGP, is to sum together a number of harmonic terms, each of which represents a partial tide. The frequencies and amplitudes of these waves can be derived from the ephemerides by constructing a time series and then filtering it, as in the Cartwright Taylor Edden (CTE) potential (Cartwright and Edden, 1973; Cartwright and Taylor, 1971), or using analytic means as for the Xi89 potential (Xi, 1987, 1989) and the potential RATGP95 (Roosbeek, 1996). Alternatively, the waves can be found by fitting coefficients to a series based on an ephemeris – the spectral method – as used in the Tamura87 and Tamura93 potentials (Tamura, 1987, 1993) and in HW95 (Hartmann and Wenzel, 1995a,b). Using the same approach, Kudryavtsev (2004) extended the number of waves to 28 000 with his new potential KSM03 and improved the time-domain accuracy to 0.025 μGal. He claims that the addition of further terms does not significantly improve this accuracy. These TGPs (except for CTE that has now been superseded), shown in **Table 3**, are those most commonly used in high-precision gravity work; they include tables (catalogs) specifying the frequencies and amplitudes of the waves that must be summed. By selecting different subsets of waves, the speed of computation can be adjusted according to the accuracy required.

Merriam (1993) compared the catalogs of Tamura (1987) and Xi (1989) with GTIDE and concluded that although their differences should be detectable using SGs, in practice, either of the catalogs could be used for SG analysis. Wenzel (1996a) compared all the aforementioned series and catalogs and concluded that the HW95 was the most accurate for high-precision work. Roosbeek (1996) noted that this is because HW95 is derived from one of the benchmark series itself and its only error should be computational. Tamura (1993) provided a harmonic development of the tidal potential that is widely used as a compromise between speed and accuracy.

3.04.2.2.2 Elastic response of the Earth

Once the tidal forcing series has been accurately computed for a precise latitude and longitude, it is necessary to allow for the deformation of the Earth to compute an approximation to the observed solid tide. This is usually done within the tide program through the use of elastic load Love numbers (h_n, k_n, and ℓ_n) for each of the harmonics (denoted by n) in the TGP. The Love numbers are computed by solving the gravitoelastic equations of motion for the Earth and finding the surface displacement $\mathbf{u}$ and potential Ψ for any kind of forced deformation (e.g., tides or loading):

$$\begin{aligned} \mathbf{u} &= h_n \Psi^1 \mathbf{e}_r / g_0 + \ell_n \nabla_1 \Psi^1 / g_0 \\ \Psi &= k_n \Psi^1 \\ \delta_n &= 1 + 2h_n - (n+1)k_n / n \end{aligned} \quad [5]$$

(e.g., Wang, 1997). Here, Ψ^1 is the perturbation in the gravity potential, $\mathbf{e}_r$ the radial unit vector, ∇_1 is the surface gradient operator, and g_0 is the surface gravity. The Love numbers completely describe any kind of deformation, elastic or inelastic, and therefore contain all the complexity of the actual Earth, that is, resonances for all the Earth's normal modes,

Table 3 Tide-generating potential used in SG data reduction

Catalog	*# Waves*	*# Coefficients*	*Degree*	*Time (nGal)*	*Frequency (nGal)*
CTE	505	1010	3	38.44	0.565
Tamura87	1200	1326	4	8.34	0.118
Xi89	2934	2934	4	6.42	0.090
Tamura93	2060	3046	4	3.08	0.046
RATGP95	6499	7202	6	2.00	0.026
HW95	12 935	19 271	6	0.14	0.002
KSM03	28 806	66 852	6	0.025	0.0004*

The last two columns refer to rms accuracies in the time and frequency domains, respectively.

*Kudryavtsev (*personal communication*, *2013*).

Modified from Wenzel H-G (1996) Accuracy assessment for tidal potential catalogues. *Bulletin d'Informations des Marees Terrestres* 124: 9394–9416; Wenzel H-G (1996) The nanogal software: Earth tide processing package ETERNA 3.30. *Bulletin d'Informations des Marees Terrestres* 124: 9425–9439, with KSM03 from Kudryavtsev (2004).

frequency dependency (e.g., Dickman, 2005). Love numbers have been frequently computed for seismic Earth models such as PREM (Dziewonski and Anderson, 1981) and given in a number of different forms; they are in principle complex numbers because of the Earth's anelasticity (e.g., Mathews, 2001). Of interest here are the real gravimetric tidal factors (δ_n and κ_n) where κ_n defines the phase of the tidal wave as observed (or computed) (note that it is 0 for an elastic model) with respect to the TGP and δ_n is found from a combination involving h_n and k_n, as in the preceding text (ℓ_n is not used in gravity). Typical elastic values are $h_n = 0.602$, 0.291, and 0.175 and $k_n = 0.298$, 0.093, and 0.043 yielding $\delta = 1.155$, 1.167, and 1.121 for $n = 2$, 3, and 4, respectively. A nominal pair of values is taken as ($\delta = 1.16$, $\kappa = 0$).

Gravimetric factors can be used to construct the synthetic tide at a station and are thus the simplest way to 'detide' a gravity record (especially a short one where tidal fitting would be problematic) and of course are perfectly suited to constructing the 'removal' signal discussed in the preceding text. They can be found empirically from a tidal analysis at a station, noting that there should be no phase lag entered into the program (e.g., Crossley and Xu, 1998). When the synthetic tide is reconstructed from the empirical gravimetric factors, both the ocean loading, considered in the next subsection, and the system phase lag will be automatically included along with the solid tide.

3.04.2.2.3 Ocean tides and loading

Ocean tides have the same frequencies as solid Earth tides, but they derive from both the vertical and (small) horizontal components of the TGP that generates water movement in shallow areas such as the continental shelves. This variable water depth loads the crust and modifies the amplitudes of the bodily tides as seen on land stations. One problem in computing ocean tides is their variability, due primarily to oceanic weather systems. A further difficulty is the need to define the coastal topography and bathymetry, particularly in remote and ice-bound locations such as Antarctica.

Ocean tidal loading is almost an invisible effect when doing tidal analysis, because the gravimetric factors are adjusted to the total tide signal that includes ocean loading. There are several problems for which ocean tides and their loading effects need to be studied in detail, and in these cases, the ocean loading must be computed separately. Depending on the location, ocean tide loading varies between 1% and 10% of the body tide.

As in atmospheric loading (see **Section 3.04.2.3**), the ocean load function is generally expressed as the convolution of a Green's function with a suitable data set representing ocean heights. Most ocean loading programs consider the heights to be given, for each wave, within cells that follow the coastlines, but the height variability due to weather systems is ignored. Ocean tide heights are now determined almost exclusively by satellite altimetry (TOPEX/POSEIDON), but this still leads to a large number of ocean tide models constructed with differing assumptions and constraints.

Ocean tide loading is an essential element of the complex problem that includes mean sea levels and global warming. It is therefore an important topic in gravimetry that has greatly benefited from the availability of high-quality SG data. Baker and Bos (2003) and Boy et al. (2003) had discussed the issue of whether SG data can discriminate between competing ocean tide models, and they emphasize the need for accurate SG calibration. Several computer programs can do the appropriate calculations. Three of the most widely used are SPOTL (Agnew, 1997; http://igppweb.ucsd.edu/~agnew/Spotl/spotlmain.html, updated in Agnew, 2012), OLFG by Scherneck (1991) (http://holt.oso.chalmers.se/loading/), and GOTIC2 by Matsumoto et al. (2001). Boy et al. (2003) showed significant differences between these programs and tide models that left the role of instrument calibration unclear. On the other hand, Bos and Baker (2005) concluded that with care, the computational errors can be resolved and the programs show almost identical results for all programs and for all tide models. This suggests that there are still phase calibration issues for SGs. They offer a new program CARGA that incorporates most of the features of the other programs.

Ocean loading corrections are frequently geared to the need for precise geodetic information in particular areas such as Antarctica (Bos et al., 2000) or the Pacific Northwest (Lambert et al., 2003).

3.04.2.2.4 Tidal analysis

Tidal analysis consists of determining the amplitudes and phases of tidal waves of specific frequencies from observational data. How many waves can be determined and to what accuracy depends on the length of the record used and on the noise characteristics of the site. In most approaches, the ocean tides are grouped in with the solid tides of the same frequency, and they cannot be separated by a simple fit of data to known tidal frequencies. Ocean tide variability is reflected in the time dependence of the gravimetric tidal factors at a particular site.

Three programs are used within the SG community for tidal analysis. The most widespread program is ETERNA, developed over many years by Wenzel (1996b) and now in its final form (Version 3.4). It consists of a suite of several programs that deal with all the common aspects of processing gravimeter data, and it can be adapted to a variety of different data sets; it also can be used for analysis of strain and tilt data as well as gravity. ETERNA is a harmonic method that does a direct least-squares fitting of the (δ, κ) factors for various wave groups to gravity data sampled at fixed intervals, for example, 1 min or 1 h. The program can handle different subblocks of data if there are gaps. The user has the choice of simultaneously fitting an admittance function to pressure or other auxiliary data. It includes polar motion and different assumptions regarding the instrument drift function. The program is well documented and available through ICET and GGP (http://www.eas.slu.edu/GGP/ETERNA34/).

The second approach is a program called BAYTAP-G that was developed by Japanese geophysicists during the 1980s (Tamura et al., 1991) and can be found at http://igppweb.ucsd.edu/~agnew/Baytap/baytap.html. The approach is a hybrid method using a combination of harmonic series and the response method (Lambert, 1974) to estimate the various components of a gravity record (i.e., the tidal parameters, a pressure perturbation effect, a smooth 'drift' function, and an irregular noise). One of the distinguishing aspects of the program is a focus on a statistical description of the drift, which in this context means the entire long-term gravity signal and not

just the instrument drift, and of the random noise. Using Bayesian estimates, the procedure involves the nonlinear estimation of a trade-off parameter between the tidal harmonic series and the residual gravity.

The third program is available, called VAV, which was described most recently by Venedikov and Viera (2004). The origins of this approach go back 50 years or more when gravimeter records were processed by the application of suitable filters of various lengths to account for the tides. Comparisons exist between VAV and ETERNA, but the former has not been as widely adopted by the SG community as the other programs. Dierks and Neumeyer (2002) compared all three programs using both synthetic data and a 1-year observed SG data set from station Sutherland (SA). They found the performance of the three programs to be similar, but with different treatments of the statistics between signals (tides, pressure, and drift) and residual gravity. The spectra of the final residuals were noticeably different between the programs, and the reader should consult the original paper for details (also for valuable tips on how to set some of the parameters in the program inputs).

A few words should be said about the response method in SG tidal analysis, as discussed by Merriam (2000). Rather than allocating for all waves within a group the same gravimetric factor, the response method selectively interpolates the gravimetric factors using only waves that seem unperturbed by noise (i.e., that are close to their theoretical expectations). Waves that seem anomalous are thus treated independently, requiring a more hands-on treatment, but one that has some practical advantages. The additional flexibility allows a greater reduction in tidal residuals because it can accommodate the more variable ocean loading.

3.04.2.3 Atmospheric Pressure Effects

The atmosphere provides a significant gravity effect (up to 10% of the tidal signal) with a transfer function (or admittance factor) that approximates $-0.3\ \mu\text{Gal hPa}^{-1}$ for a typical continental station. The effect is a combination of gravitational attraction by atmospheric density anomalies with a loading that vertically deforms the crust and mantle. For a positive atmospheric density anomaly, simple theory gives about $-0.4\ \mu\text{Gal hPa}^{-1}$ for the upward attraction and $+0.1\ \mu\text{Gal hPa}^{-1}$ for the loading. A number of well-studied empirical and physical methods exist for making a pressure correction to the gravity data, but even with the most sophisticated treatments, it is not possible to completely remove the atmospheric pressure effect.

3.04.2.3.1 Single admittance factors

Atmospheric effects on gravity became an important consideration with the higher precision and lower noise of the SG compared to previous instruments. Warburton and Goodkind (1977; henceforth called WG77) anticipated most of the issues that were revisited in many subsequent papers (e.g., Müller and Zürn, 1983; Rabbell and Zschau, 1985; Spratt, 1982; van Dam and Wahr, 1987). A useful review of the role of local, regional, and global effects was given by Merriam (1992b), from which we draw the following conclusions:

(a) Approximately 90% of the atmospheric effect comes from the local zone, defined as within 0.5° (or 50 km) of the station, and 90% of this effect is from the direct Newtonian attraction of the atmosphere; the remaining 10% is from loading and deformation of the crust.
(b) The regional zone extends from 0.5° to ~3° or from 50 to 100–500 km (depending on the topography surrounding the station), and in this zone, the atmospheric correction is small (a few %) and primarily comes from loading.
(c) The rest of the atmosphere, at distances >3°, is the global effect and contributes only a few % of the total atmospheric gravity.

Merriam found a local admittance factor of $-0.356\ \mu\text{Gal hPa}^{-1}$ and, with a variable regional correction, a combined admittance that varied between -0.27 and $-0.43\ \mu\text{Gal hPa}^{-1}$. This is consistent with the factor of -0.30 found by WG77 for frequencies <1 cpd, a factor that is widely applied for the quick 'correction' of gravity due to atmospheric pressure. Note that this factor applies to relative changes in pressure and gravity that have to be propagated from one data sample to the next by maintaining appropriate reference levels.

One way to better model the local atmosphere is to use an array of sensors located around the gravity station. Riccardi et al. (2007) reported on an experiment to improve the pressure correction for the Strasbourg station using an array of barometers at regional scales (typically a few tens of km). The improvement in this experiment, however, was found to be marginal.

3.04.2.3.2 Frequency-dependent admittance

WG77 first recognized that a single admittance factor was not appropriate at all frequencies, partly because of the separate effect of atmospheric tides at harmonics of a solar day (e.g., Elstner and Schwahn, 1997). In addition, however, the atmosphere shows more coherency toward higher frequencies (2–8 cpd) and the admittance factor from the data at a single station is larger than at lower frequencies (<1 cpd) when regional scales become important. WG77 also introduced the cross spectrum between gravity and pressure to find an admittance function in the frequency domain, and this was extended by Crossley et al. (1995) and Neumeyer (1995) who showed the advantages that could be realized under certain conditions. Frequency dependence is equivalent to allowing the admittance function to vary in time, as would arise in the passage of weather systems (e.g., Müller and Zürn, 1983).

To give some insight into these points, we show in **Table 4** a comparison of different processing methods, from Crossley et al. (1995). Five different corrections are given with the notation TD = time domain, FD = frequency domain, $\delta g(t)$ = gravity correction in the TD, $\Delta g(\omega)$ = Fourier transform of $\delta g(t)$, and $p(t)$ = pressure.

Note that fitted values of the admittance α (cases 2 and 3) are smaller than the nominal correction (case 1) because they respond to the lower coherence in the atmospheric signal at long periods and at frequencies that are harmonics of 1 cpd. This is clear in **Figure 14** that shows the amplitude component of the admittance from 2 years of gravity and pressure data from Cantley (CA). The smooth background obtained by binning segments of the spectra uses an averaging window of 1.5 cpd and represents the increasing coherence with frequency of the atmosphere as sampled from a single location. Note

Table 4 Comparison of atmospheric pressure corrections

Case	*Type*	*Formula*	*α-Value*	*rms*
1	Nominal admittance in the TD	$\delta g(t) = -0.3p(t)$	0.30	
2	Fitted admittance in the TD	$\delta g(t) = -\alpha p(t)$	0.257	1.147
3	Fitted admittance in the FD	$\Delta g(\omega) = -\alpha P(\omega)$	0.255	1.142
4	General FD admittance	$\Delta g(\omega) = -\alpha(\omega) P(\omega)$	**Figure 11**	0.672

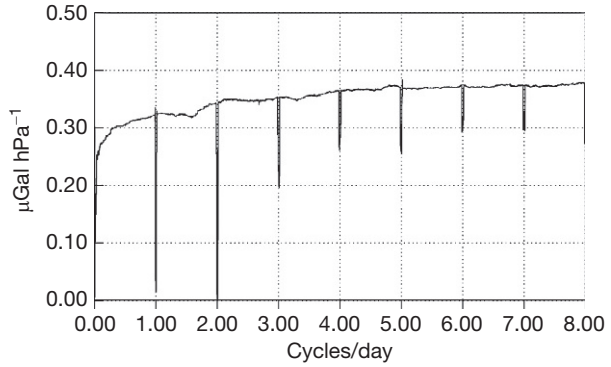


Figure 14 Amplitude of frequency-dependent barometric admittance in **Table 4**, case 4. Reproduced from Crossley D, Jensen O, and Hinderer J (1995) Effective barometric admittance and gravity residuals. *Physics of the Earth and Planetary Interiors* 90: 221–241.

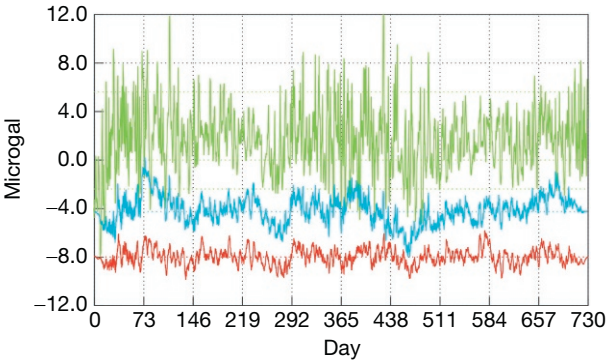


Figure 15 Pressure corrections using an empirical scalar admittance function. The upper trace (green) shows uncorrected gravity residuals (tides, polar motion, and drift removed), the middle trace is after correction using a single admittance (blue), and the lower trace (red) is after the correction using an empirical frequency-dependent admittance (traces offset for clarity). Reproduced from Crossley D, Jensen O, and Hinderer J (1995) Effective barometric admittance and gravity residuals. *Physics of the Earth and Planetary Interiors* 90: 221–241.

that the asymptotic value is consistent with the values of 0.356 - μGal hPa^{-1} quoted by Merriam. The sharp negative lines are computed separately from a narrow window of 0.01 cpd centered on the harmonics of a day and reflect the low coherence between gravity at a single location and the global atmospheric tides. There are issues with using such a function, however, because the choice of averaging window has a large effect on how much signal is removed from the gravity residuals.

In the time domain, we compare the signals in **Figure 15**. The upper trace is the uncorrected gravity, the middle trace is corrected with a single admittance (case 3), and the lower trace is for a frequency-dependent correction (case 4). As noted by Kroner and Jentzsch (1999), one should not assume that a smaller residual necessarily means a 'better' correction because the atmosphere is not the only source of gravity variations. In the period range 1–8 cpd, it is known that rainfall and hydrology also contribute significantly to gravity and this extra 'noise' considerably complicates the determination of an accurate admittance. Except for specialized studies, a frequency-dependent admittance is rarely done as part of regular processing. Aside from ease of use, a nominal value of −0.30 μGal hPa^{-1} has the advantage that such a correction can quickly be 'restored' to gravity residuals if required.

We note that the atmospheric admittance effects in coastal areas depart from continental areas (discussed in the preceding text) because water responds to surface pressure as a fluid rather than elastically as for the solid Earth. This inverted barometer effect permits the water column to adjust isostatically to ocean surface pressure, and the ocean bottom pressure does not contribute to atmospheric loading. The local admittance for near-coastal stations may therefore depart from that represented in **Figure 14**.

3.04.2.3.3 Green's functions and nonlocal pressure corrections

The admittance approach is suited to single pressure series taken at the station, but it can approximate only the local part of the atmospheric effect. For regional and global corrections, it is necessary to compute gravity directly from spatial meteorologic data. All discussions on loading start with Farrell (1972) who showed how to convolve an observed distribution of surface pressure (data) with a Green's function (or kernel) that represents the effect of a point load for a given source – station separation. The method is quite general and can be used for any kind of surface load and local and nonlocal corrections for any kind of atmospheric model.

Two-dimentional loading calculations can take one of two forms: either the surface pressure alone is used or the surface temperature is included, which allows the incorporation of thermal changes in the air column using a simple gas law (Merriam, 1992b). The former option is simpler to implement because pressure data are widely available from meteorologic centers, though sampled usually only at 6 h intervals. The thermal admittance given by Merriam is $\delta g = +0.013\ (T_c - 15°)$ μGal $°C^{-1}$, which combines the effects from local and regional zones. In the passage of a cold front, for example, with pressure and temperature changes of 3 hPa and 10 °C, respectively (Müller and Zürn, 1983), the gravity effect from temperature is only 10% (i.e., 0.1 μGal) of the pressure effect and may be reasonably ignored.

Numerical computations using global surface meteorologic data need to be done for a spherical Earth (Merriam, 1992b), so the computational task becomes nontrivial, especially when using the high-resolution pressure data. Some notable results were found by Mukai et al. (1995) and Sun (1995). Boy et al. (2002) considered the improvement to be expected at low frequencies when the regional and global loading is done explicitly using various assumptions for the vertical structure of the atmosphere and different meteorologic data sets (ECMWF (European Centre for Medium-range Weather Forecasts) at 1.125° vs. NCEP (National Center for Environmental Predictions) at 2.5°). In addition, they demonstrated that the inverted barometer assumption is appropriate for periods longer than 1 week. We show in **Figure 16(a)** and **16(b)** their results for two different corrections in zone 2 (the region $>0.5°$) that provide a reduced gravity residual compared to the local admittance. Note that the use of a simple gas law model of the atmosphere (pseudostratified loading) reduces the size of the nonlocal effect compared to the loading that uses only the surface pressure. As in Merriam (1992b), the atmospheric correction due to temperature and water vapor has been assumed to be small.

Finally, we mention the importance of including topography when making atmospheric corrections in hilly or mountainous terrain (Boy et al., 2002).

3.04.2.3.4 Three-dimensional atmospheric corrections

In recent years and of particular importance to the correction of atmospheric effects in the Gravity Recovery and Climate Experiment (GRACE) satellite data (Boy and Chao, 2005), it has been recognized that seasonal mass changes within the atmosphere are not reflected in surface pressure data. This means that an additional 3-D Newtonian attraction needs to be modeled when high-quality atmospheric corrections are required. This has been known for some time, and a correction for this effect can take the form of a simple annual sinusoidal term of amplitude about 0.8 μGal, as reported in Zerbini et al. (2001, 2002). The first complete treatment of the effect and the computations required can be found in Neumeyer et al. (2004a). Here, we show in **Figure 16(c)** sample calculations from this

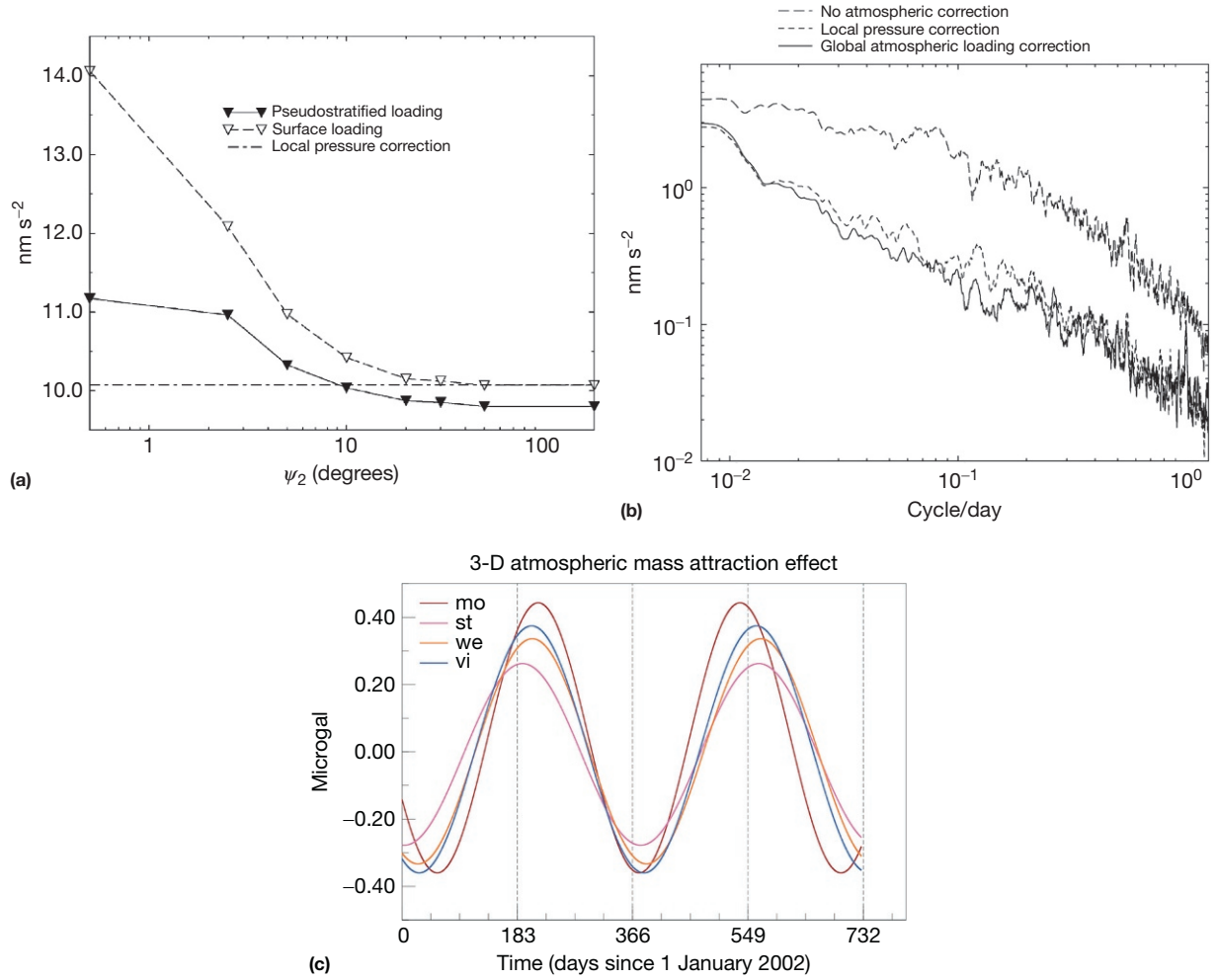


Figure 16 Atmospheric pressure corrections (a) from 0.5° to 180° using 2-D Green's functions compared to a local scalar admittance, (b) the reduction of noise in the frequency domain (panels (a) and (b) reproduced from Boy J-P, Gegout P, and Hinderer J (2001) Gravity variations and global pressure loading. *Journal of the Geodetic Society of Japan* 47(1): 267–272), and (c) seasonal air mass effect for atmospheric attraction at four GGP stations.

study for four stations in central Europe over a 2-year period (only the annual fitted function is shown). In the summer months, the air masses within the atmosphere rise and the gravitational attraction effect decreases (upward); this causes a net positive gravity effect at the ground that is about 1 μGal larger than that in winter months. Neumeyer et al. (2004a) showed that the effect is not the smooth function shown in **Figure 16(c)**, but there are sudden changes that occur over periods of days.

Further improvements have been reported by Abe et al. (2010) and Klügel and Wziontek (2009) in which different numerical approximations (e.g., 2-D vs. 3-D) were compared and, as for the simple admittance calculation, the attraction close to the station tends to dominate the effect. The annual variation is now considered to be quite small (less than a μGal because of the opposite cancelation of masses from the low and high atmospheric layers). The computational effort to include this term is not trivial, and it is being done routinely only at a few SG stations, but fortunately, there is now a service called Atmacs (Atmospheric attraction computation service), available at the website http://atmacs.bkg.bund.de/index.php. One limitation is the need to interpolate the meteorologic data that are available only at 6 h sampling and 0.5° (50 km) spacing using data sets such as ECMWF and DWD (the German weather service). This is also a factor for the data reduction in GRACE because the atmospheric pressure corrections are done on the intersatellite distance timescales that are aliased by 6 h sampling.

3.04.2.4 Calibration Issues

3.04.2.4.1 Basics

SGs are relative instruments and no 'factory calibration' is provided by the manufacturer. The instrument must be calibrated to convert its output feedback voltage, as recorded by a digitizing voltmeter, to units of acceleration. The amplitude calibration factor (GCAL) is frequently called the scale factor of the instrument and is expressed in $\mu Gal\,V^{-1}$ or $nm\,s^{-2}\,V^{-1}$, with a typical value of about $-80\ \mu Gal\,V^{-1}$ (**Table 5**). In addition, the system transfer function of the complete system (sensor and electronics) must be measured to determine the frequency dependence of both the amplitude and phase. Although it is easy to find the approximate calibration (to a few % in amplitude or phase), it is nontrivial to improve the calibration to be better than 0.1%. There are three applications for which accurate calibration is most important: (a) ocean tide loading or solid Earth tidal deformation including the FCN resonance, (b) the subtraction of a synthetic model for tides and the atmosphere whenever residuals are required to high accuracy, and (c) determining precise spectral amplitudes in normal-mode studies.

The SG output has a very large bandwidth, from ~1 s (the raw sampling) to periods as long as the data length (years to decades). It is common to refer to the long period limit as 'DC' (vaguely meaning beyond tidal periods), and some of the calibration methods effectively measure this DC amplitude calibration constant, whereas other techniques attempt to find other parts of the transfer function. Even in well-controlled laboratory experiments, it is difficult to provide a suitable acceleration at periods beyond a few hours. This is not an issue for seismometers because they are not designed to measure periods longer than 1 h. For gravimeters at tidal periods, however, one has to either extrapolate the results from shorter periods or extract the calibration from the data under normal operating conditions.

As noted by Richter (1991), some of the methods used to calibrate other types of gravimeters are inappropriate for SGs. One of these is the use of calibration 'lines' where an instrument is repeatedly moved to precise locations arranged in a horizontal or vertical line; due to the instrument size, this is

Table 5 Representative SG calibration experiments using an absolute gravimeter

Station	*Instrument*	*AG or method*	*# Drops*	*Time*	*SF (μGal V⁻¹)*	*(%)*
BH (Falk et al., 2001)	CD030_L	FG5 #101	18000	2 years	−73.690 ± 0.088	0.12
		Platform			−73.971 ± 0.023	0.03
	CD030_U	FG5 #101			−67.626 ± 0.084	0.12
		Platform			−67.922 ± 0.041	0.06
BO (Francis et al., 1998)	C024	FG5 #205	20800	9 days	−80.281 ± 0.063	0.08
		Platform			−80.341 ± 0.009	0.01
CA (Merriam et al., 2001)	T012	JILA-2	na	3 years	−78.3 ± 0.1	0.13
CB (Amalvict et al., 2001b)	C031	FG5 #206	15778	6 days	−76.098 ± 0.169	0.22
			46560	12 days	−75.920 ± 0.061	0.08
MA (Imanishi et al., 2002)	T011	FG5 #210	100000	27 days	−92.801 ± 0.034	0.04
					−92.851 ± 0.049A	0.06
					−92.879 ± 0.036B	0.04
MB (Francis, 1997)	C021	FG5 #202	275468	47 days	−78.457 ± 0.001	0.06
MC (Falk et al., 2001)	C023	FG5 #101, 103, 206	18000	4 years	−74.822 ± 0.137	0.18
		Platform			−74.824 ± 0.013	0.02
ST (Hinderer et al., 1991)	TT05	JILA-5	5600	1 day	−76.05 ± 0.55	0.72
ST (Amalvict et al., 2001a)	C026	FG5 #206	412244	3 years	−79.19 ± 0.05	0.06
ST (Amalvict et al., 2002)	C026	FG5 #206	450000	4 years	−79.40 ± 0.03	0.04
SY (Iwano et al., 2003)	T016	FG5 #203	55743	15 days	−58.168 ± 0.061	0.10

Scale factors (SF) are by direct regression except (A) tidal analysis and (B) modified least squares. Station codes BH, etc. are given in Crossley (2004).

impractical for the SG. The application of electromagnetic or electrostatic forces is also difficult due to the need for a precise knowledge of the reference force.

For a relative calibration, an SG is compared over a period of time with another instrument whose calibration is known (including calibration by an AG). In this case, the two sets of measurements can be compared indirectly, for example, by comparing the elastic gravimetric factors from a tidal analysis or by a direct regression of the data from one instrument to the other. For absolute calibration, two methods have been used – either a test mass is moved with respect to the gravimeter and the Newtonian attraction is measured directly or the instrument itself is subject to a known acceleration provided by an acceleration platform.

3.04.2.4.2 Amplitude calibration and relative methods

3.04.2.4.2.1 Calibration using a theoretical tide

The best theoretical solid Earth tidal models are accurate to about 1 nGal in frequency-domain amplitude. This is about 10^{-5} of the full tidal amplitude (300 μGal), so it would seem a straightforward task to calibrate an instrument using the tidal signal. In practice, however, the observed tidal amplitudes are a combination of theoretical amplitudes, elastic response of the Earth, and ocean tidal loading. Variability in ocean crustal loading of about 1 μGal limits the accuracy of tidal calibration to about 0.3% accuracy.

Several authors have argued that tidal analysis of an SG record is a useful calibration tool (e.g., Goodkind, 1996). Of equal importance is the relative stability of the amplitude calibration, and Merriam (1995a,b) had inferred this for the model TT70 at Cantley using the tidal admittances of the M_2, O_1, and K_1 waves from 3 years of data. He concluded that the amplitude calibration was stable at the 0.013% level and the phase calibration to within 0.01° for M_2. Similar analyses of long duration (several years) of tidal observations suggest indeed reliability in calibration factors at the level of 0.01–0.07% (Rosat et al., 2009a,b; Sato et al., 2004). The use of a theoretical tide (solid tide + elastic yielding + ocean tide loading) is still model-based and therefore of a different character to the other methods listed in the succeeding text.

3.04.2.4.2.2 Calibration using AGs

For several reasons, the use of an AG has become the most widely used method in recent years. First, the SG is not disturbed during the calibration because the AG measurements are done on a nearby pier or an adjoining room. The instrument separation needs to be kept small, however, to prevent horizontal gradients due to local gravity variations (e.g., hydrology), from affecting the instruments differently. Second, even though the AG has a repeatability of 1–2 μGal over a typical measurement campaign, with care, a precision of about 0.05% can be achieved (**Table 5**).

Another advantage is that the processing of the data from the two instruments is straightforward, and no complicated physical effects have to be accounted for because the instruments measure the same temporal changes whatever their origin. Finally, the collocation of the two instruments also enables the long-term drift of the SG to be estimated, and this provides obvious advantages in cross-checking the AG measurements.

Hinderer et al. (1991) reported an early calibration of a model TT70 SG using a JILA-5 AG in Strasbourg. They introduced the direct regression technique whereby all SG and AG are used without correction, save for allowing different linear drifts between the two lasers of the JILA instrument. When tested against theoretical tides for O_1 (a common method for European stations because the solid tide O_1 is large and the ocean tide loading is small), agreement was found at about the 0.1% level. The full-size T012 at Cantley was calibrated a number of times using AGs. Bower et al. (1991) obtained 0.4% from tidal gravimetric factors using a JILA-2 and Merriam et al. (2001) reported on a series of nine calibrations to achieve a precision of 0.13%.

Several authors have engaged in extensive campaigns to intercompare the instruments. Francis (1997) reported on 47 days of recording involving a large number of AG drops. A precision of about 0.1% was achieved both by linear regression and by comparing tidal amplitude factors. A significant finding of this study was that the precision of calibration was not significantly improved for times longer than 5 days, suggesting that this is an optimum period for such a comparison. Similar results were found for the C024 instrument in Boulder (Francis et al., 1998) that agreed very well at a level of 0.08% with the absolute calibration using an acceleration platform.

Okubo et al. (1997) performed tests using FG5 #107 at two of the Japanese SG sites, T007 in Esashi and T011 in Matsushiro. Their results pointed out that the standard deviation of the SG set values coincident with the AG measurements was 0.3 and 0.62 μGal for Esashi and Matsushiro, respectively; this is better than the error in the AG residuals at the two sites (1.55 and 1.52 μGal).

A sampling of other results of AG calibrations is given in **Table 5**. Amalvict et al. (2001a, 2002) showed a series of 28 different calibration (totaling almost half a million drops) runs over a 4-year period for ST, with an average length of 5 days each. The regression method gave an overall precision of 0.04% with a repeatability of 0.1%. Imanishi et al. (2002) presented a study of the various ways to cross calibrate the AG and SG using a simple regression, tidal factors, and a combined method (**Table 5**). They found that the SG is not the only instrument subject to 'drift' and the AG showed a bias of several μGal over the 1 month of the experiment. Nevertheless, they were able to extract a very precise calibration of about 0.04%, which approaches the best calibrations by an acceleration platform method. Finally, Falk et al. (2001) and Harnisch et al. (2002) had given useful results on a number of SGs and AGs, together with some acceleration platform calibration results.

Note that for BH, the two spheres have different calibration factors – this is true of all dual-sphere instruments. Also, MC is one of many stations that have been visited by several different AGs. The 'errors' are generally the precision of a least-squares fit and are known to underestimate uncertainties obtained from differing sets of measurements. Thus, some SGs have been DC-calibrated to better than 0.05%, but others are probably closer to 0.15%.

Francis and Hendrickx (2001) turned the calibration issue around and used an SG as the reference signal in the calibration of a spring gravimeter. This is possible due to the extremely low drift and high precision of the SGs and produces

good results, especially when the variable drift of the spring gravimeter is modeled along with the calibration, as in Meurers (2002).

3.04.2.4.2.3 Calibration using spring gravimeters

Many early SG calibrations came from the side-by-side comparison of data from spring gravimeters that had already been calibrated either in the laboratory or on calibration lines. Due to limitations in the original calibrations, it is unlikely that the amplitude factors can be trusted below about the 0.5% level (Richter, 1991).

A recent calibration experiment of SG C026 in Strasbourg (France) by means of two FG5 AGs and new-generation spring meters (Scintrex CG-3 M and CG5 and Micro-g LaCoste gPhone) was recently done by Riccardi et al. (2012). This study confirms that, owing to the time variability of their sensitivity, such spring meters, even if well calibrated, cannot be used as a stable reference for SGs and will not replace AGs for precise SG calibration. Therefore, currently, the method using parallel recording with absolute gravity meters is still the most practical and reliable calibration approach for SGs. The study by Meurers (2012) also involved an SG, AG, and the potential of a spring gravimeter to contribute to SG calibration.

3.04.2.4.3 Amplitude calibration and absolute methods

3.04.2.4.3.1 Calibration using a moving mass

Warburton et al. (1975) first described the use of a heavy-mass calibration experiment. They rolled a 321 kg hollow steel sphere filled with mercury to a fixed position under the SG, thus generating a 10 μGal test signal. Within the accuracy of the various parameters, the most critical being the exact distance between the center of the test sphere and the niobium sphere, they found a precision of 0.2%. In such an experiment, other factors limiting accuracy are the homogeneity of the sphere, the exact geometry of the position of the sphere with respect to the niobium sphere, and loading and tilting effects that are induced by the heavy mass itself.

A similar calibration was done by Goodkind et al. (1993) who used a moving mass system to test the inverse square law of Newtonian gravity. A spherical mass was moved vertically under the SG in a specially designed chamber; an accuracy of 0.09% was achieved for two different types of sphere material. A limit in this type of calibration is the uncertainty in the gravitational constant, which at 0.01% is still smaller than the precision of the above calibrations.

Achilli et al. (1995) implemented a unique moving mass system for the full-sized T015 SG in Brasimone as part of a test for constancy of the gravitational constant. Their ring, weighing 272 kg, is raised and lowered around the SG from strong supports in the roof; but a limited range of about 1 m reduced the mass effect to 6.7 μGal. During two calibrations, they found a precision of 0.2% and a repeatability of 0.3%.

3.04.2.4.3.2 Calibration using a moving platform

The idea of using an acceleration platform for the calibration of spring gravimeters was used successfully for LCR gravimeters by Van Ruymbeke (1989). The idea was applied to SGs by Richter (1991) who pioneered the development of the Frankfurt calibration platform. The instrument design, in which the dewar and cold head are mounted separately, limits the maximum vertical range through which the gravimeter can be lifted. For a sinusoidal signal of amplitude 5 mm, Table 6 shows the maximum accelerations that are possible. Note that for periods longer than 2 h, the amplitude is smaller than atmospheric and hydrologic effects. Other limitation to this technique arises from the gravity gradient effect, the mechanical stresses deforming the support systems, and the need to avoid tilting the gravimeter, which causes the horizontal resonance of the sphere to be excited.

Table 6 Maximum accelerations available from a moving platform, amplitude ±5 mm

Period (s)	*Acceleration (μGal)*	*Output (V)*
300	219.30	3.96042
480	85.60	1.50698
600	54.81	0.95092
900	24.36	0.43222
1200	13.70	0.25497
2400	3.42	0.07782
3600	1.52	0.02585
7200	0.38	0.00065

Modified from Richter B (1991) Calibration of superconducting gravimeters. In: *Proceedings of the Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Europeen de Geodynamique et de Seismologie, 3, Luxembourg*, pp. 99–107.

This type of calibration also allows the phase response to be measured through the phase offset between the instrument response and the acceleration function (Richter et al., 1995a). A calibration precision of better than 0.02% in amplitude and 0.1 s in time delay has been achieved using this system (Table 6).

3.04.2.4.4 Phase calibration

3.04.2.4.4.1 Response of the analog filter

The ideal amplitude gain of the electronic analog antialiasing filter should be constant from DC to the corner frequency of the antialiasing filter, and the ideal phase response should be linear with frequency. These aspects are generally met in most SG recordings but depart from the ideal at high frequencies, which for gravity data analysis means periods shorter than about 1 h. The GGP filter, used on the GWR gravity card, is an 8-pole Bessel filter designed to achieve the desired linearity.

The phase response at longer periods can be well represented by $\phi(\omega) = -\alpha\omega$, where α is a constant, ϕ is measured in degrees, and ω is in cycles per day (cpd). Given $\phi(\omega)$, it is simple to find the group delay $\tau(\omega) = -\mathrm{d}\phi/\mathrm{d}\omega$, which is usually measured in seconds. For a linear phase filter, one has (with α in degrees per cpd)

$$\tau(\omega) = 240\alpha = \text{constant} \qquad [6]$$

For example, for the IDA filter, $\alpha = 0.15$ deg cpd^{-1} so $\tau = 36.0$ s. For the GGP filter board at 1 s sampling, $\alpha = 0.035$ deg cpd^{-1} so $\tau = 8.4$ s. In order to usefully measure the nonelastic response of the Earth to tidal waves, accuracies of 0.002° are required in the tidal wave phases (Wenzel, 1994). The phase lag of the recording system therefore has to be

known at least to the same accuracy, that is, 0.5 s. An electronic calibration method is necessary to achieve this precision.

3.04.2.4.4.2 System response

The antialiasing filter is not the only component in the data acquisition system that causes time delays and phase shifts. They can occur throughout the electronics, including the feedback sensor, the digitizing voltmeter, and the time stamping of the signal onto the hard disk. Therefore, the overall phase response is best determined *in situ* and not assumed from the design characteristics of only the antialiasing filter.

3.04.2.4.4.3 The step response method

The amplitude and phase response (the transfer function) of the complete system can be measured by injecting a known voltage to the feedback coil and recording the output voltage of the feedback network. The transfer function completely describes the electronic response of the system but is not a substitute for determining the instrument calibration, that is, the conversion from voltage to acceleration. The injected voltage can be either a step function, which gives both amplitude and time responses but is sensitive to noise, or a series of sine waves that suppresses noise but requires a longer measurement time (Van Camp et al., 2000). The third possibility is to inject a randomized input signal to the instrument and find the transfer function by cross spectral analysis, but Richter and Wenzel (1991) concluded that this method is best suited to instruments with electrostatic or electromagnetic feedback, such as the spring gravimeters used in the IDA network.

The initial experiment to determine instrumental phase lag was by Richter and Wenzel (1991) in which they outlined the method and gave results for several spring gravimeters and one SG, the TT60 at Wettzell. Results for the latter showed an average time lag of 38.73 ± 0.14 s and a phase determination of 0.002° for diurnal tidal waves, at the target accuracy level noted in the preceding text. Wenzel (1995) later determined even better results for an LCR meter with a group delay accuracy of ± 0.004 s, far superior to the nominal value determined from the electrical components themselves.

Van Camp et al. (2000) had discussed in most detail the phase response measurement for a compact SG, in this case the C012 at Membach. The electronic step was applied to various combinations of filter type and time intervals (1–4 min), and the output was Fourier-analyzed to determine the amplitude and phase response. The time lag results showed an accuracy of ± 0.003 for the GGP1 filter and ± 0.075 s for the raw gravity signal output. The sine wave method gave very similar results, confirming the integrity of the method. The amplitude gain of the electronics system was determined to be flat from 500 to 2000 s, the longest period determined by either method. Further details are given on the GGP website. A recent study by Francis et al. (2011) based on the comparison of the transfer functions of three SGs operating in Walferdange (Luxembourg), Membach (Belgium), and Pecný (Czech Republic) shows that the transfer functions cannot be assumed to be the same as that computed with a theoretical mode. This means that a precise phase determination should be performed for each SG.

3.04.2.5 Other Corrections to Residual Gravity

3.04.2.5.1 Polar motion

One important signal contained in an SG record is the 14-month (435-day) oscillation of the rotation pole or the Chandler wobble. Richter's (1983) first observation of this signal, with only 5 μGal amplitude, was a turning point in the refinement of gravity residuals. Since then, every SG station has recorded data that when suitably processed show the polar motion (e.g., Harnisch and Harnisch, 2006a). Naturally, some records are very clear and others not so clear, depending on the epoch (amplitude of the motion) and the quality of the instrument and site.

With most data sets, it is not difficult to see the polar motion once tides and atmospheric pressure are subtracted. Indeed, for the most part, the highly accurate space geodetic series for the polar orientation that is given on the IERS website can be used directly at any of the gravity stations. A simple conversion is usually made between the (x, y) amplitudes of polar motion (m_1, m_2) in radians and the gravity effect δg in μGal:

$$\delta g = 3.90 \times 10^{-9} \sin 2\theta [m_1 \cos\lambda - m_2 \sin\lambda] \qquad [7]$$

(e.g., Wahr, 1985) where (θ, λ) are station latitude and longitude. The numerical factor includes the nominal value of 1.16 for the gravimetric delta factor, whereas fitted solutions are usually closer to $\delta = 1.18$ (e.g., Harnisch and Harnisch, 2006b; Loyer et al., 1999). In many studies, the polar motion is considered as a quantity very accurately determined from space geodetic data, but, as we show later, the measurement of the gravity effect allows us to find the long-period transfer function of the Earth.

3.04.2.5.2 Instrument drift

From a historical viewpoint, particularly for spatial gravity surveys, the term 'drift' has been applied to almost any unwanted time-varying gravity signal. This usage persists in the SG literature and sometimes causes confusion. Here, we use drift as an instrument characteristic, because all other gravity variations have specific causes.

From an instrumental point of view (e.g., Goodkind, 1999), drift is likely to be either a linear or exponential function of time, but its size is not easy to predict. The exponential behavior can be reset after a loss of levitation or other magnetic changes within the sensor. Under normal operation, the user can generally assume that drift will level off after installation and gradually become more linear as time progresses. Representative values of instrument drift are frequently < 4 μGal year^{-1} where these have been checked carefully with AGs (see the preceding text).

From a processing point of view, other functions such as polynomials have been used as a model for instrument drift, but there is no physical reason to prefer such a choice. We recommend that an exponential drift be assumed from some initialization event for the instrument and, later, this may be replaced with a simple linear function (Van Camp et al., 2013). Drift is not to be confused with a secular change of gravity, even though the two cannot be separated except using combined SG–AG observations. Note that when using BAYTAP-G tidal analysis, the

drift signal is not just that of the instrument but includes long-term effects (Tamura et al., 1991).

3.04.2.5.3 Hydrology

Hydrology is perhaps the most complex of the intermediate scale (hour–year) variations in gravity (e.g., Harnisch and Harnisch, 2006a). This is due to two factors. The first is its variability, due largely to the local water storage balance at the station that involves many components (rain and snowfall, soil moisture, evapotranspiration (ET), and run-off). Rainfall is relatively easy to assess, as is the groundwater level, which is usually measured in a nearby well. The direct measurement of soil moisture is not easy and has not been done at most SG stations.

The second problem is one of length scales. The connectedness (permeability) of the soils and groundwater system is inhomogeneous at the local length scales (m to km), and so, an assessment of the amount of moisture surrounding an SG involves extensive measurements. This quantity predominantly affects the attraction term rather than the loading term.

As a result, even though groundwater variations are extremely useful, they are not entirely reliable for the purpose of determining an admittance factor for 'correcting' gravity residuals. The simplest case is a horizontal layer (Bouguer slab) of moisture of thickness h and fractional porosity ϕ, yielding a gravity perturbation:

$$\delta g = 2\pi G\rho\phi h = 0.42\phi\,\mu\mathrm{Gal\,cm^{-1}} \qquad [8]$$

(e.g., Crossley et al., 1998). This is similar in application to the single admittance factor for atmospheric pressure, but the porosity (and permeability) can usually be estimated only very approximately, thus limiting the accuracy of the correction. Adding to the complications are SG stations located underground where a soil moisture layer may be present both above and below the instrument. In this case, the hydrology correction becomes quite problematic and is usually treated as a research problem rather than a correction that is part of normal processing. An exception is the paper by Meurers et al. (2007) in which they demonstrated that making hydrologic corrections can improve the detection of offsets and the precision of tidal analysis. A new GGP service for atmospheric, ocean, and hydrological loading can be found at http://loading.u-strasbg.fr/GGP/.

3.04.2.5.4 Residual gravity

Having completed the aforementioned processing steps and corrections, one finally arrives at a residual gravity series that represents the unknowns including effects such as ocean currents, secular changes in elevation due to tectonics, gravity changes due to slow and silent earthquakes, or eigenmodes of the Earth (perhaps from the Earth's core). In addition, of course, all the subtle effects of mismodeling will be applied, which may include wrong assumptions regarding disturbances, offsets, drift, loading, and hydrology.

On one conclusion, there can be little doubt – the ability of SGs to reliably measure effects at the 0.1 μGal level has opened up many interesting scientific possibilities and posed many challenging issues that will be discussed in the next sections of this chapter.

3.04.3 Scientific Achievements Using SGs

3.04.3.1 The Global Geodynamics Project

The GGP (http://www.eas.slu.edu/GGP/ggphome.html) is an international research effort that was launched as a SEDI (Study of the Earth's Interior) initiative at the Vienna IUGG General Assembly in 1991. In 2003, it changed to become IAG (International Association of Geodesy) Inter-Commission Project and reports to Commission 2 (The Gravity Field) and Commission 3 (Earth Rotation and Geodynamics). It consists of a worldwide network of SGs, currently about 30 instruments, run by independent national groups. The groups agreed to provide vertical gravity acceleration data in a standard form, basically untouched raw data decimated to 1 min samples and sent at the end of every month to a database (http://ggp.gfz-potsdam.de). As indicated in **Figure 17**, GGP stations are sparsely distributed worldwide, with only two regional clusters of instruments, one in Europe and a smaller one in Japan. The coverage in the southern hemisphere is still weak, despite the effort of installing stations in Australia, Indonesia, and Antarctica by Japanese colleagues, in South Africa by GFZ Potsdam (Germany), and in South America by BKG (Germany) (TIGO project). Phase I of the project was the period 1997–2003 and phase II the period 2003–07; currently, GGP is a continuing endeavor associated with the International Association of Geodesy.

The scientific objectives of the GGP cover geophysical phenomena throughout the wide period range of the instruments (from 1 s to several years), covering topics such as normal modes, mantle rheology, tides, solid Earth–oceans–atmosphere interactions, hydrology, and Earth rotation. **Figure 18** represents schematically the gravity spectrum that is observable by SGs ranging from seconds (ocean noise) to several years (secular changes). We refer the reader to the EOS article by Crossley et al. (1999) where a full description is provided. Other review papers on SGs have appeared by Goodkind (1999), Hinderer and Crossley (2000), Meurers (2001a), Hinderer and Crossley (2004), and Neumeyer (2010). The last review on the GGP network and the underlying scientific challenges is by Crossley and Hinderer (2009).

We will show in the succeeding text some of the most interesting results, which owe their existence to the collection of the worldwide GGP data of high quality. Many results in this section are quoted in their original units of nm s^{-2} (0.1 μGal).

The wide spectrum of geophysical phenomena that are observable with SGs is evident in **Figure 18**. Basically, the range of observable periods (or characteristic time constants) covers eight orders of magnitude from 1 s to several years. The lowest period detectable by SGs is ~1 s, because of the feedback system limitation, and on the left, the figure shows background noise mainly caused by ocean noise with two dominant peaks at 5–6 and 10–15 s. At slightly longer periods, we have the seismology region including the normal modes generated by earthquakes – periods up to 54 min that are the gravest period of the Earth elastic eigenmodes. Between 150 and 500 s (2–7 mHz), these modes form the incessant oscillations ('hum') unrelated to earthquakes but rather of atmospheric and/or oceanic origin. At periods longer than about 6 h (depending on the core stability profile), another

class of eigenmodes are the gravity–inertial modes (also called core modes) predominantly confined to the liquid core. A particularly interesting and isolated long-period oscillation is the Slichter mode (actually a triplet) arising from the translation of the solid inner core. Its period, between 4 and 8 h, depends primarily on the density jump of the inner core boundary (ICB).

From 4 h to 18.6 years, there are many spectral lines due to lunisolar tides, the most important of which are semidiurnal and diurnal with amplitudes exceeding several tens of μGal for the largest ones. The study of diurnal gravity tides includes a resonance effect due to the existence of wobble modes for an Earth model with a fluid outer core (FCN, free core nutation) and a solid inner core (FICN, free inner core nutation). Toward

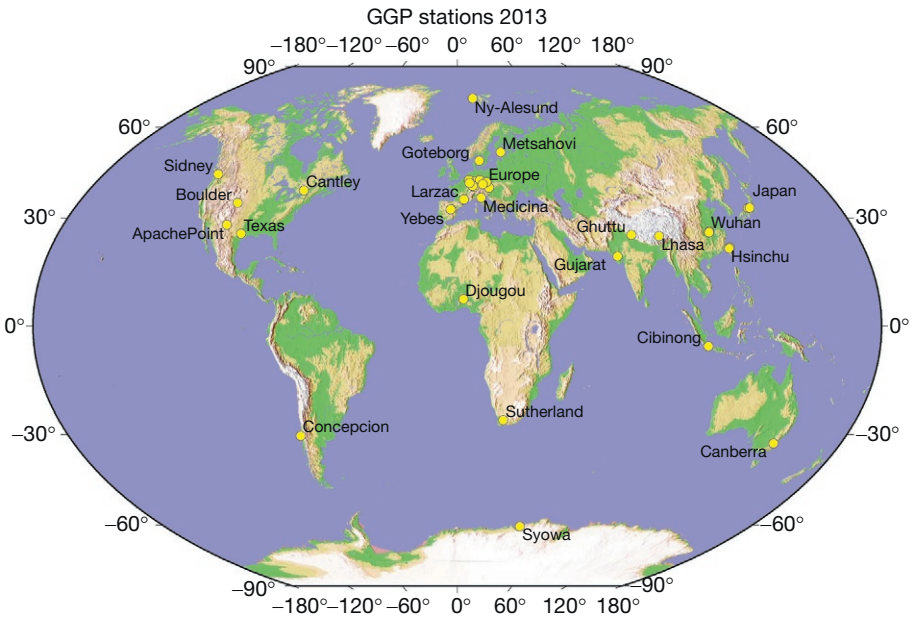


Figure 17 Geographic location of the stations of the GGP (Global Geodynamics Project) in 2013. Please note that status with open data access is still provisional for some new stations like that in Sidney.

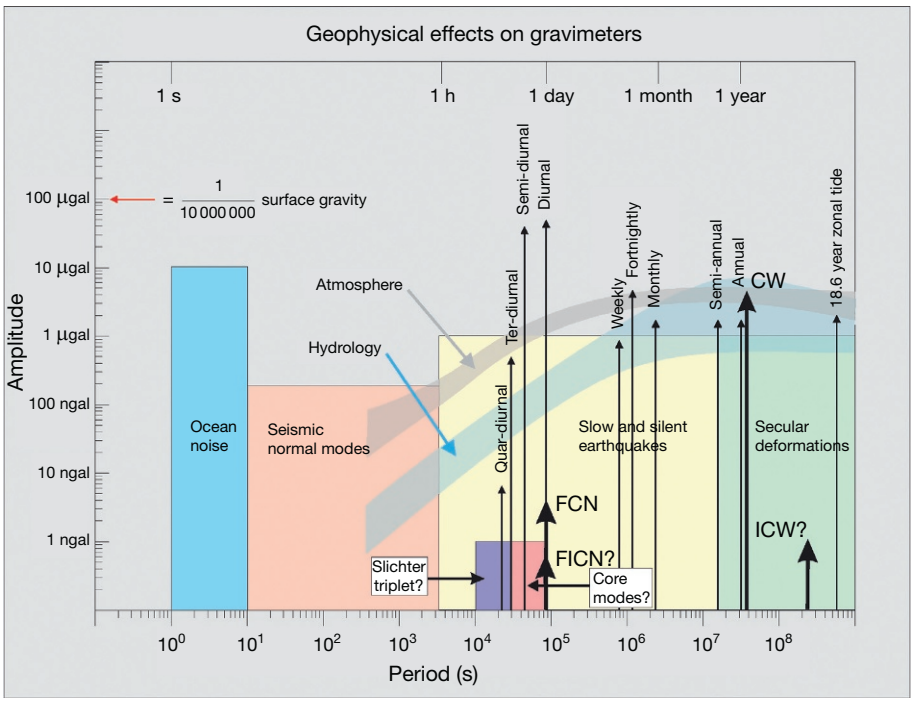


Figure 18 Characteristics of geophysical phenomena observable by SGs.

longer periods, there are many signals (atmospheric pressure, snow, soil water content, ocean circulation, and Earth rotation) that are related to the seasonal solar cycle, which lead to contribute to a strong annual term in gravity. CW stands for Chandler wobble, which is another rotational mode of the Earth with a 435-day period that causes polar motion (Section 3.04.2.5.1); its observation provides valuable information on mantle rheology at long period. The ICW, or inner core wobble, has not been observed but is predicted to have a period of several years with unknown excitation (see Section 3.04.3.8).

On the right of the figure, we have long-period secular deformations, for instance, gravity changes caused by postglacial rebound (PGR) or tectonics. Also indicated in shaded bands are the effects due to atmospheric pressure and hydrology. In contrast to periodic phenomena that have sharp spectral lines, these effects are broadband and cover a large part of the spectrum, atmospheric effects being roughly one order of magnitude larger than hydrologic effects. Figure 18 is a normalized amplitude spectrum where a sinusoidal wave with unit amplitude will always appear with unit amplitude regardless of the period of observation. On this type of spectrum, noise would decrease as a function of record length. We have chosen to represent the typical variability of hydrology and atmospheric pressure-induced effects in gravity (equivalent time standard deviation) as time-independent. The two bands are derived from the analysis of water table level and air pressure records over several years available at some of the GGP stations.

Figure 19 (left) depicts the gravity changes due to solid Earth tides and ocean tidal loading in Strasbourg for January 2010. The solid Earth tides (mostly the beating between semidiurnal and diurnal waves) can reach 2500 nm s^{-2} peak to peak and the ocean loading 50 nm s^{-2} peak to peak. An example of the different long-period (>10 days) contributions to the time-variable gravity signal is shown in Figure 19 (right) for the same station. The largest contribution still remains the solid Earth tides with several hundreds of nm s^{-2} and multiple periods (fortnightly, monthly, semiannual, and annual). Ocean loading is by far smaller of the order of several nm s^{-2} (this term can be larger for a station near the oceans). Nontidal ocean loading is slightly larger and exhibits a strong seasonal feature. Atmospheric loading is broadband and can easily cause gravity changes of several tens of nm s^{-2}. Hydrology (soil moisture + snow) is predominantly seasonal and can also lead to gravity perturbations of similar amplitude. Finally, the Earth's rotation changes (polar motion and length of day) induce a gravity signal of several tens of nm s^{-2} with a dominant beating between an annual term and the Chandler wobble.

We now review some scientific studies in the various period ranges as discussed in Figure 18. Section 3.04.3.2 deals with the short-period seismic phenomena and other normal modes up to a 1-day period. In Section 3.04.3.3, atmospheric effects on gravity are described, while tidal contributions (fluid core resonance effect and linear and nonlinear ocean loading) are discussed in Section 3.04.3.4. Section 3.04.3.5 deals with nontidal ocean circulation, hydrology contributions are covered in Section 3.04.3.6, and ice melting effects are discussed in Section 3.04.3.7. Section 3.04.3.8 is on Earth rotation and polar motion effects. Section 3.04.3.9 deals with tectonics, and Section 3.04.3.10 considers the problem of the calibration/validation of gravity satellite data with SG ground observations. Finally, Section 3.04.3.11 suggests future possibilities using superconducting gravimetry.

3.04.3.2 Seismic and Subseismic Signals

Investigation with SGs of the seismic normal modes excited by large earthquakes has led to some new and impressive results, due primarily to the low noise levels of SGs at periods longer than 500 s as indicated in Figure 20 (Rosat et al., 2003a). At periods longer than 1000 s, the best SGs have lower noise than the worldwide seismometer limit NLNM (new low-noise model, Peterson, 1993); noisier SG stations cross the NLNM at much larger periods (for instance, 4.2 h for the station BE). This extensive compilation of all the GGP stations extends earlier results from just a few SG stations but which already were convincing in terms of low noise (Banka and Crossley, 1999; Van Camp, 1999). In addition, metrological comparisons between SGs and broadband seismometers (Freybourger et al., 1997; Hinderer et al., 2002b) and between AGs and SGs over a large spectral range (Crossley et al., 2001; Francis et al., 2004) have demonstrated the excellent characteristics of SGs. Since new instruments appeared recently in the GGP network, an updated comparison of noise levels has been done by Rosat and Hinderer (2011). This study shows that the most recent observatory SG types do not show any further improvement with respect to the compact models operating at the same stations. At BFO in Germany, the experience of the dual-sphere OSG with a lower sphere heavier than usual has shown that the instrumental and site conditions make this station the least noisy one in the GGP network at frequencies higher than 0.1 mHz (see Figure 9).

Of recent large earthquakes, the 2001 Peru event of magnitude $M = 8.4$ strongly excited the long-period seismic modes, and significant observations were made by the GGP network. In particular, the fully split ${}_0S_2$ multiplet (with five individual singlets) that has been rarely visible on a single instrument could be fully analyzed from the Strasbourg C026 instrument (Rosat et al., 2003a) but was also present at other SG stations. A recent paper by Häfner and Widmer-Schnidrig (2012) on ${}_0S_2$ excited by different large earthquakes and observed on the GGP network confirms the specific contributions of SG measurements to the determination of 3-D density distribution inside the Earth. We note in passing that SGs frequently saturate on large earthquakes, but this is not a problem for analyzing normal modes.

One important new result was the detection of the overtone ${}_2S_1$ (see Figure 21), which is an elastic mode, unlike the Slichter triplet ${}_1S_1$ whose period is determined primarily by a gravitational restoring force. The eigenfunctions of the two modes share some similarities (Rosat et al., 2003b). The detection of this mode strongly benefited from the stacking method proposed by Courtier et al. (2000), and also used by Guo et al. (2006), that is applicable to all degree 1 modes with their associated surface gravity changes.

The $M_w = 9.3$ Sumatra huge earthquake on 26 December 2004 even more strongly excited the low-frequency seismic modes and provided a unique opportunity to improve the determination of the period and Q of the gravest seismic

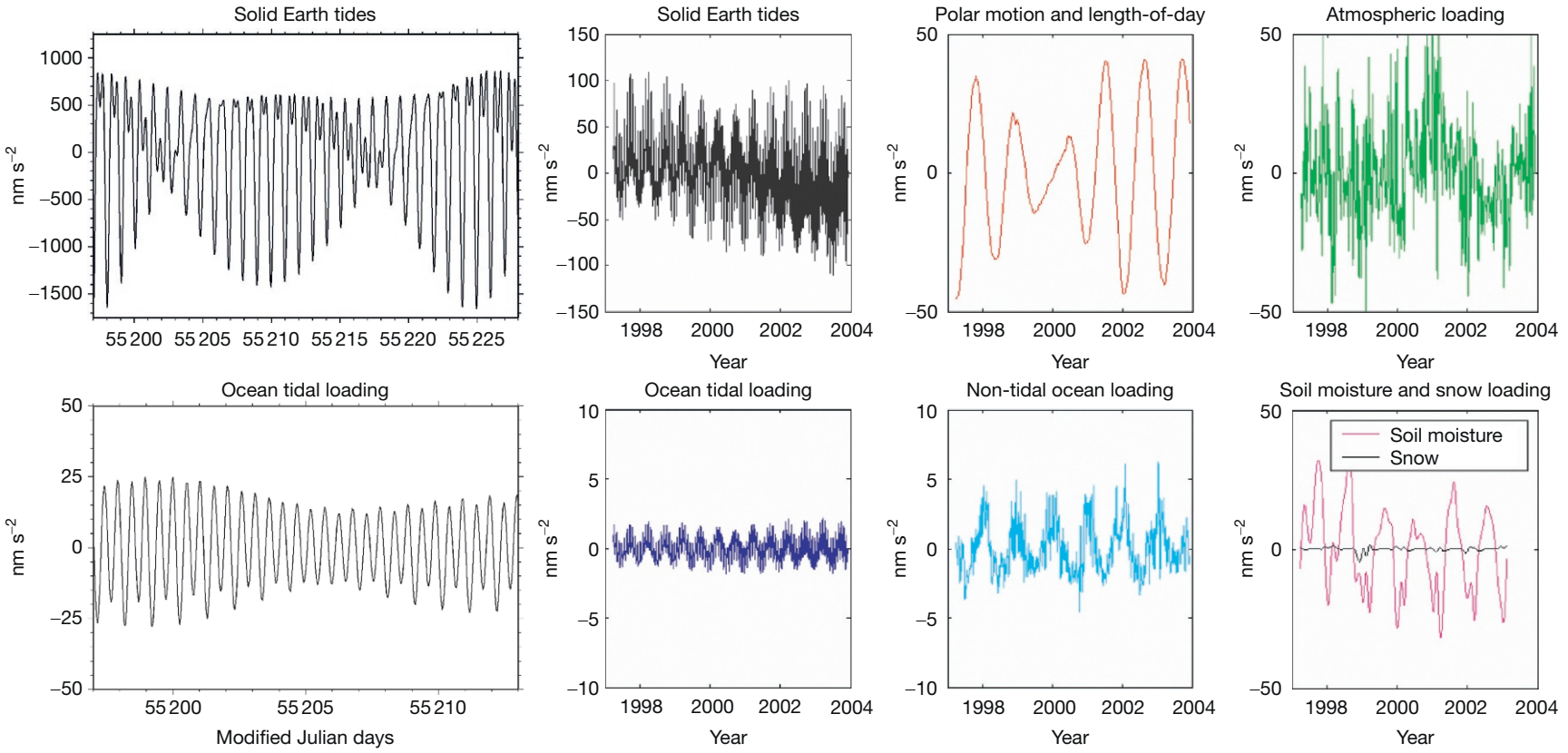


Figure 19 Different contributions to time-varying gravity (in nm s^{-2}) in Strasbourg. Left is for the solid Earth tides and tidal ocean loading during January 2010; right is for the long-period (period >10 days) part in solid Earth tides, ocean tidal and nontidal loading, polar motion and length of day changes, atmospheric, soil moisture and snow loading. Courtesy of J.-P. Boy.

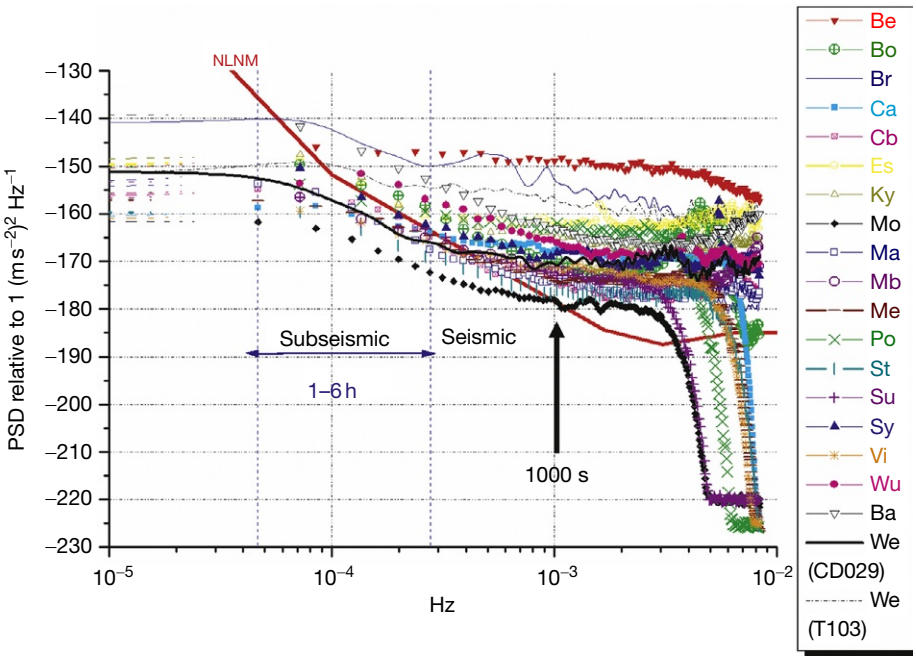


Figure 20 Noise levels of the SGs from the GGP network. Adapted from Rosat S, Hinderer J, Crossley D, and Rivera L (2003) The search for the Slichter Mode: Comparison of noise levels of superconducting gravimeters and investigation of a stacking method. *Physics of the Earth and Planetary Interiors* 140: 183–202.

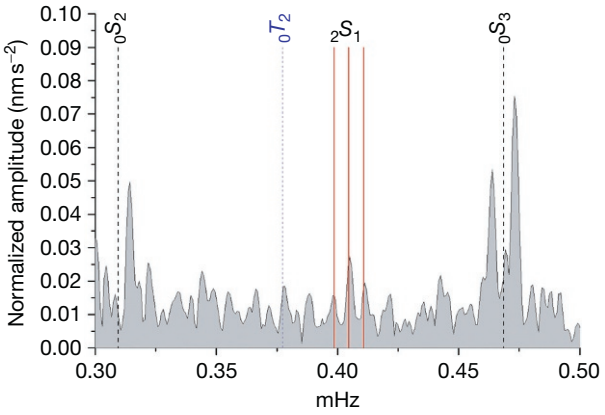


Figure 21 Normalized amplitude spectrum (in nm s^{-2}) in the frequency band 0.3–0.5 mHz after the $M_w = 8.4$ Peru event exhibiting the triplet of the $_2S_1$ mode. Reproduced from Rosat S, Rogister Y, Crossley D, and Hinderer J (2006) A search for the Slichter triplet with superconducting gravimeters: Impact on the density jump at the inner core boundary. *Journal of Geodynamics* 41: 296–306.

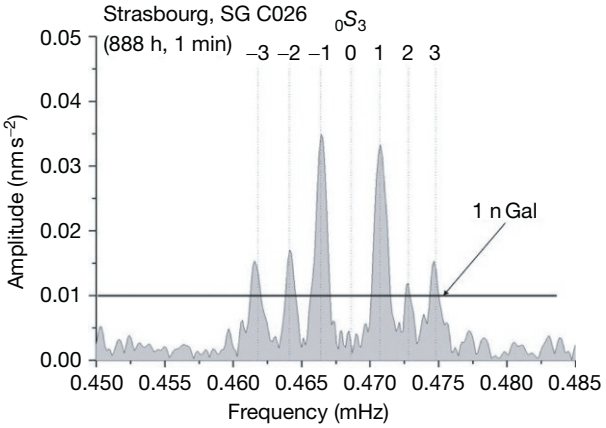


Figure 22 Observation of the well-resolved singlets of $_0S_3$ at Strasbourg using 888 h of data after the 2004 Sumatra earthquake of magnitude $M = 9.3$. The singlet $m = 0$ was not excited at Strasbourg. Adapted from Rosat S, Sato T, Imanishi Y, et al. (2005) High-resolution analysis of the gravest seismic normal modes after the 2004 $M_w = 9$ Sumatra earthquake using superconducting gravimeter data. *Geophysical Research Letters* 32: L13304, http://dx.doi.org/10.1029/2005GL023128.

modes. An example of the strong signal to noise ratio is shown in the succeeding text in **Figure 22** where all singlets of the spheroidal mode $_0S_3$ are visible on the Strasbourg SG (Rosat et al., 2005). Xu et al. (2008) used 18 recordings of the normal modes from the 2004 Sumatra earthquake to focus on the $_0S_0$ radial mode. They show that the amplitude and Q from SGs have a significantly smaller variance than from seismometers. Unfortunately, this study could not verify that the geographic distribution of observed SG amplitudes confirms the theoretical predictions proposed by Rosat et al. (2007), thus requiring further studies.

We also show the background normal modes of the Earth (frequently referred to as the 'hum') that was first discovered in the Syowa, Antarctica, SG record (Nawa et al., 1998) and later seen on other SG records (Nawa et al., 2000; Rosat et al., 2003b; Widmer-Schnidrig, 2003). **Figure 23** shows the PSD in the frequency band 1.5–5.5 mHz of the gravity data observed in Strasbourg during the period 1997–2001 with SG C026. The spectral peaks of the fundamental spheroidal modes are clearly visible. Rather than selecting periods without

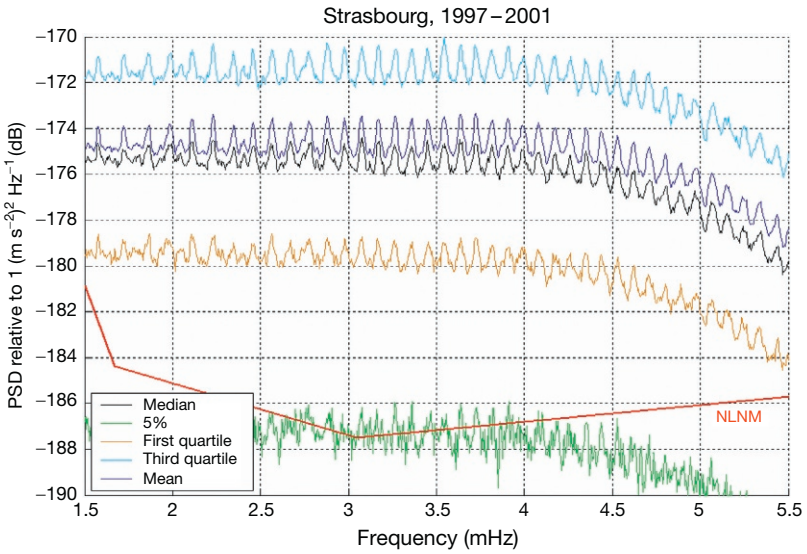


Figure 23 Power spectral density of the gravity data observed in Strasbourg during the period 1997–2001 clearly exhibiting the presence of the 'hum.' Courtesy of S. Rosat.

earthquakes above a specific magnitude, we used a statistical approach by computing the quartiles of all available gravity data in period of study.

The discovery of the hum has generated numerous studies in observational seismology and also led to early theoretical arguments, suggesting that the preferred excitation mechanism is not earthquakes, but rather the atmosphere (Kobayashi and Nishida, 1998; Lognonne et al., 1998; Nishida and Kobayashi, 1999; Nishida et al., 2000, 2002; Roult and Crawford, 2000; Suda et al., 1998; Tanimoto and Um, 1999; Tanimoto et al., 1998). Two array-based studies in seismology, however, suggest an alternative explanation in terms of interactions between the atmosphere, the oceans, and the sea floor in case of a stormy weather (Rhie and Romanowicz, 2004; Tanimoto, 2005). Recently, Traer et al. (2012) used USArray data (STS-2) to provide more evidence in support of the oceanic excitation hypothesis.

One major goal of GGP was to detect the translational motion of the solid inner core (the Slichter triplet $_1S_1$); its period(s) would bring a new constraint on the density contrast at the ICB (Rosat et al., 2006) and possibly also on viscosity just above the ICB (Smylie et al., 2001). We will not review here previous studies and controversies on the theoretical issues in modeling this eigenmode nor different claims for detection and counter studies; instead, we refer the reader to review in Hinderer et al. (1995) and Rosat et al. (2006).

It is, however, worth showing the current status of the search. For example, in **Figure 24** from Rosat et al. (2004), we see that there is obviously no clear observational evidence in this stacking of SG data of any of the theoretically predicted triplets. Evidently, many spectral peaks may emerge slightly above the background level of 0.01 nm s^{-2}, but as pointed out by Florsch et al. (1995a), these are probably only statistical fluctuations. Some claims have been made (e.g., Guo et al., 2006), but they are (as yet) generally unsupported by other observations and not believed by the SG community.

Although not yet observed, the identification of $_1S_1$ remains a unique topic for the GGP network, and interest in this topic remains very high. This has led to the development of techniques for stacking the SG records, all based on the original idea of Cummins et al. (1991) as developed in Courtier et al. (2000). Despite low noise levels in the band from 3 to 6 h and despite the methodological attempts to enhance any global signature of degree 1 triplet (see Rosat et al., 2003a), the problem probably remains one of excitation, as discussed by Crossley (1993). Even for the largest modern earthquake (Chile 1960, magnitude 9.5), the seismic moment yields a gravity signal of no more than 1 nGal, and to be observed, this amplitude needs to be maintained for some time beyond the earthquake. An alternative possibility, first used by Smylie (1992), is that the excitation could be intermittent or randomly caused perhaps by fluid motions in the outer core. Rosat and Rogister (2012) computed the excitation effect on the translational motion of the inner core due to various causes: atmospheric load, fluid core pressure, earthquakes, explosions, and finally the impact of a meteoroid. It turns out that fluid pressure acting at the core boundaries at timescales shorter than the Slichter eigenperiods would be the best candidate.

The possibility that SGs can contribute to large-scale tests of gravitational theory by monitoring gravitational frame effects through either the translational motion of the Earth (essentially the 6 h Slichter triplet) or the $_0S_0$ and $_0S_2$ long-period seismic modes at high precision has been raised by Shiomi (2008a,b). Limitations on the identification of such weak effects at the required level of precision, however, rest on our ability (or lack thereof) to model and almost completely remove competing environmental signals from ocean tidal loading and atmospheric loading.

3.04.3.3 Atmospheric Loading

As discussed earlier, several methods of correcting for the atmosphere have been proposed. We now show a specific example of the 2-D atmospheric pressure corrections done by

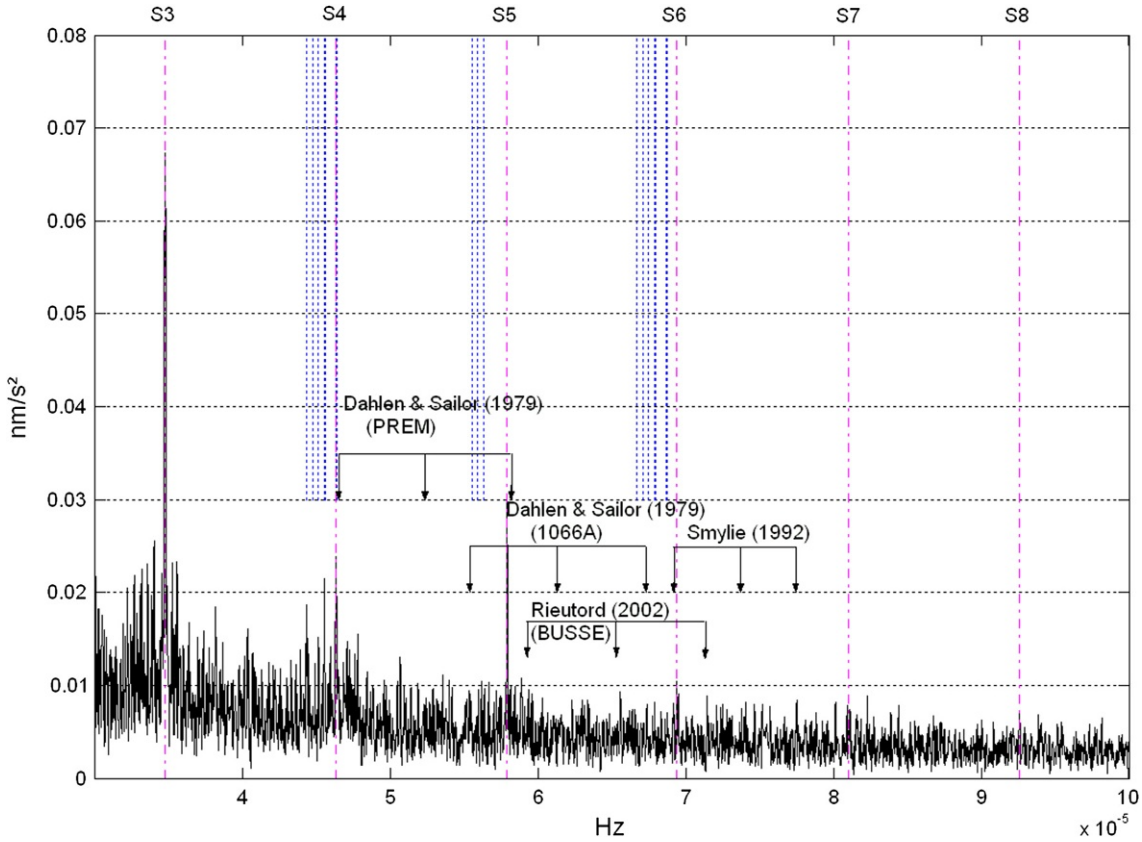


Figure 24 Product spectrum resulting from the multistation analysis method applied to five SGs. The three frequencies computed according to the splitting formula of Dahlen and Sailor (1979) for the Earth models 1066A and PREM, predicted by Rieutord (2002) based on the Earth's model by Busse (1974), and the peaks that Smylie (1992) identified in SG observations to be the inner core translation are indicated by vertical arrows. The modulations of the diurnal harmonic S_1 of the atmosphere from S_3 to S_8 (vertical dashed and dotted lines) and nonlinear tides (vertical dotted lines) around 4 and 6 h are also shown. Reproduced from Rosat S, Hinderer J, Crossley D, and Boy J-P (2004) Performance of superconducting gravimeters from long-period seismology to tides. *Journal of Geodynamics* 38: 461–476.

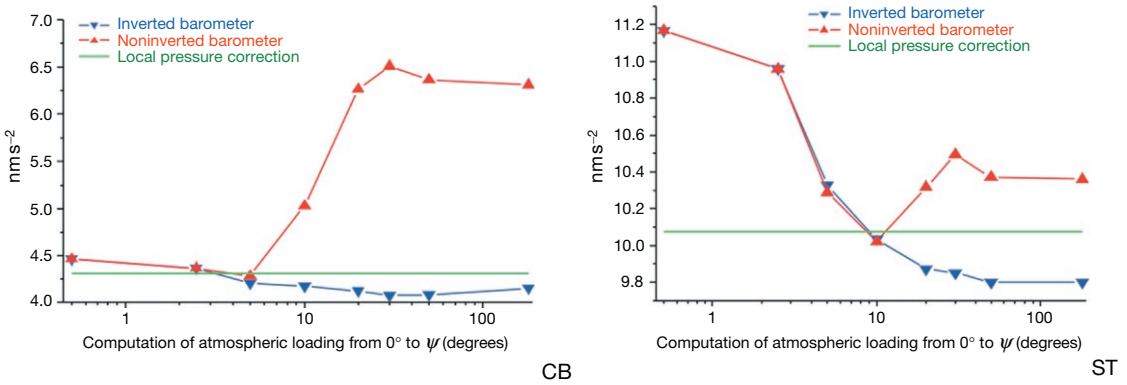


Figure 25 Variance reduction of gravity residuals (in nm s^{-2}) due to different pressure loading corrections (2-D load with inverted or noninverted barometer response, local admittance); left is for Canberra (CB) station in Australia and right for Strasbourg (ST) station in France. Reproduced from Boy J-P, Gegout P, and Hinderer J (2001) Gravity variations and global pressure loading. *Journal of the Geodetic Society of Japan* 47(1): 267–272.

Boy et al. (1998, 2001, 2002) and other groups in Japan (Mukai et al., 1995), using 2-D pressure data originating from the weather prediction centers such as NCEP or ECMWF. Figure 25 shows the computations for the stations Canberra in Australia and Strasbourg in France, where the residual gravity level (in nm s^{-2}) is plotted as a function of the solid angle in degrees around the station.

Except for small angles (close to the station), the 2-D correction performs better than the local barometric admittance close to -3 nm s^{-2} hPa^{-1} (straight horizontal line). Because of

the known correlation between length and timescales in the atmospheric processes (Green, 1999), long-period pressure effects (exceeding a few days) require large-scale loading computations to be adequately represented. The results from a coastal station of the GGP network (CB station in **Figure 25**) clearly discriminated against the use of the NIB (noninverted barometer) ocean reaction to air pressure changes.

As we previously discussed, in order to have the most accurate atmospheric corrections for the new gravity satellite missions, CHAMP (Challenging Minisatellite Payload) and GRACE, 2-D pressure loading computations have been extended to 3-D modeling where pressure, temperature, and humidity data varying in height are also taken into account in addition to the surface data (see Boy and Chao, 2005; Neumeyer et al., 2004a; Svenson and Wahr, 2002). It was recently shown for the station Medicina with the help of balloon radio sounding that there is a seasonal vertical air mass change in the atmosphere without ground pressure changes, which is caused by warming and cooling (Simon, 2002) and which leads by attraction to a gravity effect of the order of 1 μGal. A recent study by Abe et al. (2010) reviews the differences in the atmospheric gravity reductions due to the use of different models of 2-D or 3-D atmospheric models.

In an attempt not only to model the atmosphere as a source of noise but also to investigate actual meteorologic effects in gravity, Meurers (1999, 2000, 2001b) analyzed SG measurements from the Vienna station and found an interesting phenomenon. It appears that during some weather disturbances, rainfall occurs without generating a large ground pressure change but causing a significant gravity drop. Meurers (1999) model suggests that vertical convective air motion (air mass exchange or water transport) does not alter the ground pressure (total air column mass unchanged) but does modify gravity through Newtonian attraction. This is an example where gravity could be of indirect use to meteorologists to indicate air movements without detectable ground pressure signature. A more detailed study on the gravity changes following heavy rains was done by Meurers et al. (2007) suggesting that, in the case of large thunderstorms, gravitational effects are caused not only by the rain but also by meteorologic processes like vertical convection.

3.04.3.4 Tides and Nearly Diurnal Earth Wobbles

As discussed previously, Earth and ocean tides are by far the largest components of surface gravity changes. The SGs have brought two areas of improvement to tidal studies. First, the high sensitivity of these meters enables them to detect small-amplitude tidal signals previously hidden in noise (e.g., nonlinear ocean tides) and to retrieve with better precision larger tidal signals (see Ducarme et al., 2002). A clear example of the high signal to noise ratio of the gravity data from the GGP network is given by Ducarme (2011) who could retrieve tidal waves generated by the third-degree tidal potential despite their low amplitude; the lack of ocean tide models deriving from the third degree is a primary concern to go further for selecting the optimal body tide model.

With this high precision, Xu et al. (2004a) revisited the question of the possible latitude dependence of tidal gravimetric factors. Using 19 GGP stations, they found that the discrepancy of the four principal waves (O_1 and K_1 in the diurnal band and M_2 and S_2 in the semidiurnal band) between observations and theoretical models (Dehant et al., 1999; Mathews, 2001) is $<0.2\%$. This means that there is no significant latitude dependence. Second, the much lower instrumental drift of SGs versus mechanical spring meters permits more precise studies of long-period tides (M_f, M_m, S_{Sa}, and S_a) (Boy et al., 2006; Ducarme et al., 2004; Hinderer et al., 1998; Mukai et al., 2001; Sato et al., 1997a).

3.04.3.4.1 Resonance effects in diurnal tides

Of the tides themselves, the largest components occur in both semidiurnal and diurnal frequency bands. Within the diurnal band, some waves are affected by a resonance that occurs due to the nearly diurnal free wobble (NDFW), also called FCN, which is a differential rotation of the fluid outer core with respect to the mantle. In a corotating reference frame, the FCN period is of course nearly diurnal; in an absolute frame, it is $\sim$430 days. The observation of this resonance requires precise amplitude and phase measurements of the diurnal tidal waves that are close in frequency to the eigenfrequency. In particular, the small-amplitude waves ψ_1 and φ_1 are critical in the retrieval of the FCN parameters, that is, the period and damping of this resonance mode. A very clear example of this can be found in Hinderer et al. (2000) where there is a noticeable improvement in the FCN adjustment when using data from the compact SG (C026) compared with using earlier data from the TT70 in Strasbourg (see their **Figures 4** and **6**).

We do not discuss here the numerous papers using GGP data to derive the FCN parameters and refer the reader to previous reviews on this subject (Hinderer and Crossley, 2000, 2004). A study combining the analysis of six SGs by Sun et al. (2002) leads to an FCN period of 429.0 (424.3–433.7) sidereal days and a Q factor of 9500 (6400–18 700). A more global analysis by Xu et al. (2004a) who used simultaneously tidal gravity observations at 19 GGP stations leads to an FCN period of 429.9 (427.2–432.7) sidereal days and a Q value about 20 000 (12 000–72 000). The FCN period is in good agreement with space geodetic studies (Defraigne et al., 1994; Herring et al., 1986; Merriam, 1994; Neuberg et al., 1987). However, in most of the earlier studies, the gravity-derived Q factor was much smaller than that obtained using VLBI (Herring et al., 1986). An artifact of the analysis is sometimes even a negative Q (see Table 1 in Florsch and Hinderer, 2000; Sun et al., 2002, 2004). The Q discrepancy between gravimeters and VLBI measurements is not due to imprecision in the gravity observations, but is rather of a methodological nature. The classical least-squares method usually applied to the determination of the resonance parameters (obviously a linearized form applied to a nonlinear problem) is inadequate because it implies Gaussian statistics (Sato et al., 2002; Sun et al., 2004) that are not correct for this problem. Florsch and Hinderer (2000) were the first to demonstrate that an appropriate Bayesian method is required to solve for the FCN parameters due to the nonstandard form of the probability distribution for Q.

Figure 26 shows the Bayesian approach applied to Canberra (CB) SG data by plotting the joint probability distribution for the eigenperiod T and quality factor Q. Taking a vertical slice through the distribution shows that the shape is

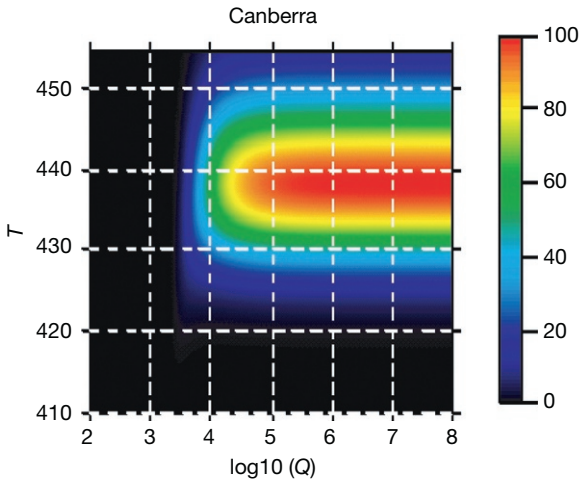


Figure 26 Joint probability density function for eigenperiod T and quality factor Q using SG data from the Canberra (CB) station. Courtesy of N. Florsch.

somewhat Gaussian for the period T, leading to values found in previous studies based on the least-squares method. This is no longer the case for a horizontal slice in which the distribution of Q shows a preferred range of high values exceeding 10^4 (including infinity that means no damping at all) in agreement with estimates from lunisolar nutation observations by VLBI.

The probabilistic Bayesian approach was compared to the classical method using the least-squares method based on the Levenberg–Marquardt algorithm by Rosat et al. (2009a). If different parameters are used (e.g., inverting $\log(Q)$ instead of Q), the two approaches lead to similar results; the eigenperiod and Q factor derived from tidal gravity agree then with those obtained from VLBI nutation data.

3.04.3.4.2 Ocean loading

3.04.3.4.2.1 Semidiurnal and diurnal tides

We have previously discussed the ocean loading correction to gravity and referred to the articles of Baker and Bos (2003), Boy et al. (2003), and Bos and Baker (2005). Clearly, observations from the GGP network can be compared to computations from different global ocean tide models, some of them being partly hydrodynamic, including the often quoted classical Schwiderski (1980) model, or assimilating data derived directly from satellite altimetry.

The present situation, from the comprehensive treatment by Bos and Baker (2005), is that the class of recent ocean tide models, specifically GOT00 (Ray, 1999), NAO99b (Matsumoto et al., 2000), and FES99 (Lefevre et al., 2002), is in better agreement with SG observations than earlier models such as Schwiderski (1980). When differences are carefully examined, most models are in agreement and there is no 'best' model valid in all areas of the world. One must mention that the spread along the real (in-phase) axis of the tidal residuals is in general larger than along the imaginary axis (out-of-phase) related to possible amplitude calibration problems. By contrast, the out-of-phase component seems to be more reliable because of a better determination of the instrumental phase lag (Van Camp et al., 2000) and is hence a strong validation tool for ocean tides.

A corollary to this investigation is the limits we can place on Earth's inelasticity from the gravimetric amplitude factors and phase delays after correction for ocean tidal loading. This was pointed out by Baker and Bos (2003) where they compared for some waves the gravimetric amplitude factor and phase delay to some reference elastic or slightly inelastic values (Dehant et al., 1999). In particular, they used the small value (≤ 0.2 μGal) of ocean loading for wave O_1 in Europe for testing inelasticity in the Earth's tidal response with European SG stations. These observations could help in rejecting inelasticity models exhibiting an increase in amplitude larger than 0.3%. The small phase lag of a few hundredths of a degree is consistent with the Mathews (2001) inelastic body tide model for 0_1.

3.04.3.4.2.2 Long period tides

We have pointed out that the SGs have a very small instrumental drift compared to classical spring meters, and this is why the investigation of long-period tides is particularly suited to SG data. In a recent paper, Boy et al. (2006) analyzed long series from 18 GGP stations to estimate the ocean loading for the monthly (M_m) and fortnightly (M_f) tides. The available models were an equilibrium tide (Agnew and Farrell, 1978), Schwiderski (1980), and three recent hydrodynamic models with satellite altimeter data assimilation NAO99b (Matsumoto et al., 2000), FES99 (Lefevre et al., 2002), and TPXO.6 (Egbert and Erofeeva, 2002). They concluded that the uncertainty for M_m is still too large to discriminate between the newer models. On the other hand, the hydrodynamic models for M_f are clearly closer to the SG observations than the equilibrium tidal model or the older model proposed by Schwiderski (1980) (see **Figure 27**).

3.04.3.4.2.3 Nonlinear tides

We have seen that the response method (which is classical in oceanography) is widely used in tidal analysis of SG records (e.g., Lambert, 1974). Using this method, Merriam (1995a,b, 2000) showed that nonlinear tides can be seen in SG records. At CA (Cantley) in Canada, it is possible to see clearly these small abnormal waves originating from nonlinearities in the ocean tidal response in the Bay of Fundy. The sensitivity inferred from the SG measurements is equivalent to 1 mm of open ocean tide, which means that these observations are a unique tool for validating these small ocean tidal waves that exist only near the coastlines (e.g., Sinha and Pingree, 1997).

Some years ago, Florsch et al. (1995b) identified small signals of a few nGal in the SG residuals with periods around 6 h from Strasbourg and Cantley stations. A more systematic study by Boy et al. (2004) on various GGP stations demonstrated that these signals are indeed due to nonlinear tides in the quardiurnal frequency band.

Figure 28 shows the fair agreement for the M_4 nonlinear tide in both amplitude and phase between SG observations and the predicted contributions from recent ocean tidal models (Flather, 1976, Mog2D and Pingree and Griffiths, 1980, 1981). Mog2D (2-D gravity wave model) is a barotropic, nonlinear model from Lynch and Gray (1979) that was later developed for tidal- and atmospheric-driven applications at both coastal and global scales.

Due to the integrative properties of the gravity loading Green's functions, inland SG observations act as a

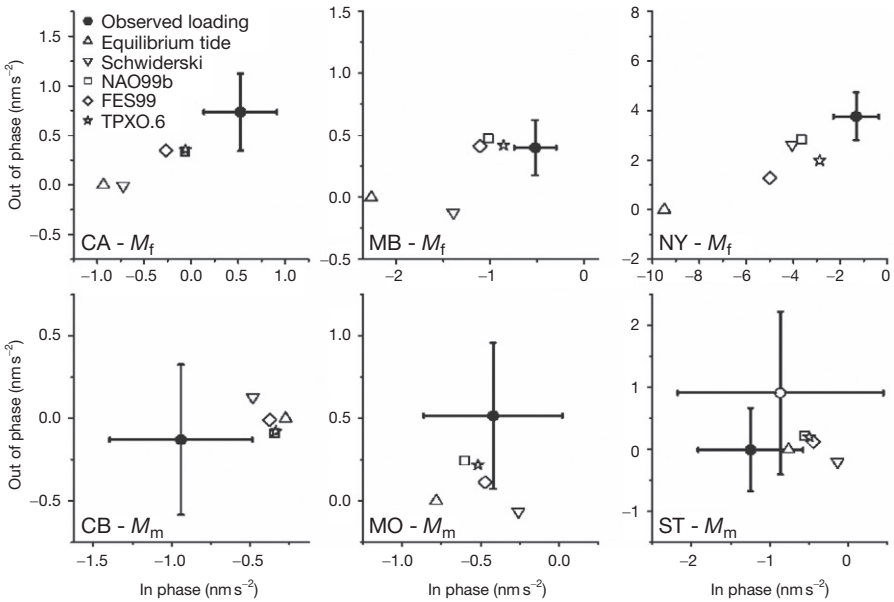


Figure 27 Observed tidal loading for M_f (top) and M_m (bottom) waves and estimated tidal loading for six SGs and five ocean models. Reproduced from Boy J-P, Llubes M, Ray R, Hinderer J, and Florsch N (2006) Validation of long-period oceanic tidal models with superconducting gravimeters. *Journal of Geodynamics* 41: 112–118.

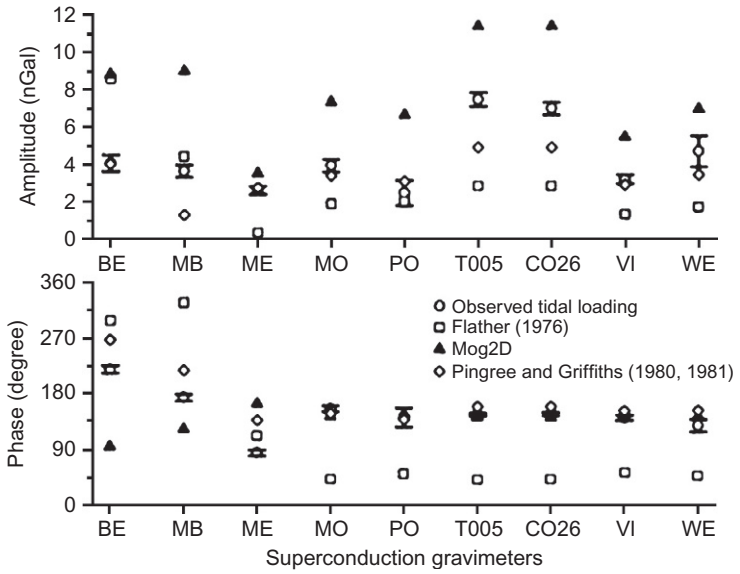


Figure 28 Amplitude (nGal) and phase (in degree with respect to Greenwich) of M_4 observed tidal loading and loading predictions according to recent tidal ocean models. Reproduced from Boy J-P, Llubes M, Ray R, et al. (2004) Non-linear oceanic tides observed by superconducting gravimeters in Europe. *Journal of Geodynamics* 38: 391–405.

complementary large-scale validation tool to point-like tide gauge observations of nonlinear ocean tides. The latter are generated only in shallow water near the coasts and hence mostly escape detection by satellite altimeters like TOPEX/POSEIDON or JASON.

3.04.3.5 Nontidal Ocean Circulation

In addition to tidal oceanic contributions, nontidal effects related to the general oceanic circulation are also detectable in SG measurements. Virtanen and Mäkinen (2003) investigated the loading effect of the Baltic Sea on the Metsähovi (Finland) instrument (T020), which is located only 15 km away from the open sea. They found a good correlation between SG residuals and sea-level changes from the nearby Helsinki tide gauge. Short-period variations are mostly driven by wind stress moving water only locally, whereas long-term variations are caused by water exchange through the Danish Straits. It is therefore useful to combine gravity observations with tide gauge measurements (and with

precise positioning measurements) to better test the loading from Baltic Sea.

Sato et al. (2001) made another important contribution to nontidal effects using SG records from Esashi, Canberra, and Syowa to investigate the nontidal annual contribution from sea-level changes (see also Fukuda and Sato, 1997). They demonstrated the importance of the steric correction to sea surface height (SSH) change – this is the coefficient used to compensate the thermal expansion of the oceans due to sea surface temperature change. The steric part does not involve any additional mass change and hence does not alter gravity by loading; thus, inland SG measurements are a unique tool to distinguish between steric and nonsteric SSH components.

Fratepietro et al. (2006) were the first to show in the Membach SG data (200 km from the coast) the loading gravity signature of storm surges on the European shelf. Another investigation of this storm surge effect was done by Boy and Liard (2008) who were able to show a better agreement between model and observations when performing a global atmospheric loading correction rather than just a local pressure correction.

Kroner et al. (2009) investigated in SG records the seasonal effects of nontidal oceanic mass shifts by using different ocean circulation models. The models predict effects of the order of 12–27 nm s^{-2} peak to peak, while the SG observations indicate a seasonal change of about 10 nm s^{-2} at distances of several hundreds of km away from the coast.

3.04.3.6 Hydrology

Early studies in gravity focused mainly on tides and oceanic and atmospheric effects – starting with the references cited above by Warburton and colleagues in the 1970s. In the last decade, however, more and more attention has been paid to the hydrology signature in gravity, especially in SG data where the drifts are typically linear and only a few μGal per year. The earlier studies are restricted to modeling local effects by trying to find correlations between gravity residuals and a relevant hydrologic parameter such as water table level, rainfall, or soil moisture (see, e.g., Crossley and Xu, 1998; Harnisch and Harnisch, 2002; Ijpelaar et al., 2002; Imanishi et al., 2006; Kroner, 2001; Kroner et al., 2004; Longuevergne et al., 2009; Meurers et al., 2007; Takemoto et al., 2002; Virtanen, 2001). A typical data set over a 2-week period is shown in **Figure 29**. Most authors use an admittance (in μGal per mm of water) that depends on local porosity and permeability (neither of which is well characterized by point measurements). Another difficulty is in separating local effects from a regional or even continental hydrology signal (see van Dam et al., 2001a,b) knowing that both will share a similar seasonal variation because the meteorologic forcing is similar.

In the study by Goodkind and Young (1991), rainfall was the cause of a gradual 70 μGal gravity increase at geothermal complex The Geysers in California over a period of 60 days that recovered after only 3 weeks. It is more usual, however, to find a rapid increase of gravity with rainfall and much slower recovery, as in the case of heavy rains at Table Mountain Gravity Observatory in Boulder CO during the Spring of 1995 (Crossley and Xu, 1998; Crossley et al., 1998; Van Dam and Francis, 1998). The SG responded to the groundwater rise with a recharge time constant of 4 h and a discharge time constant

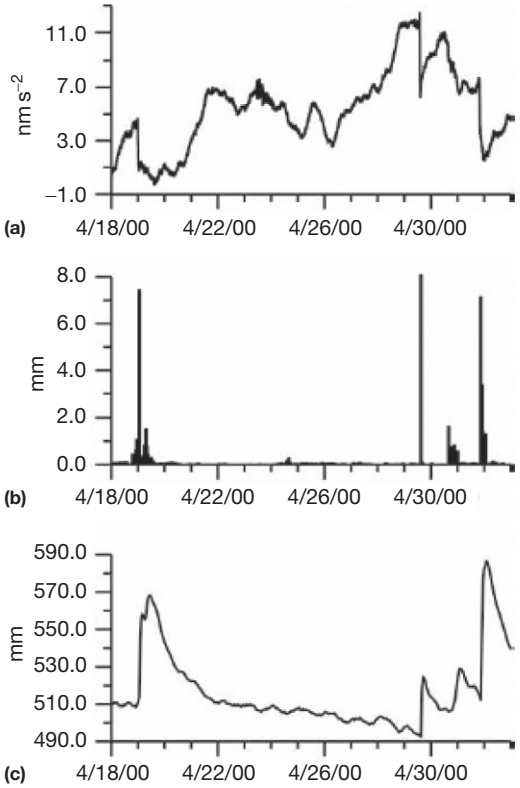


Figure 29 Gravity residual observations (a), accumulated precipitation (per hour) (b) and groundwater table level (c) at Moxa, Germany, from 18 April to 2 May 2000. Reproduced from Kroner C, Jahr T, and Jentzsch G (2004) Results from 44 months of observations with a superconducting gravimeter at Moxa/Germany. *Journal of Geodynamics* 38(3–5): 263–280.

of 92 days, a pattern that has been confirmed at a number of other sites but with different time constants (e.g., Harnisch and Harnisch, 2006a).

An early study using soil moisture was done by Peter et al. (1995) who found gravity changes of a few μGal from variations in soil moisture between 23% and 32% coupled with a significant groundwater effect. Similarly, Bower and Courtier (1998) used SG, groundwater, precipitation, and ET together to model gravity variations of 10–20 μGal over annual timescales that accounted for 90% of the residual variance in the residual gravity. Heavy rainfall was also a feature of the Bandung SG site in Indonesia, where soil moisture variations accounted for 80% of the short-term gravity changes (Abe et al., 2006). Rainfall effects have also been observed in conjunction with coseismic gravity changes (Nawa et al., 2009).

At a number of SG sites, there has been a large effort to characterize the hydrologic cycle using both fixed and moveable gravimeters as well as hydrologic instrumentation. For instance, Van Camp et al. (2006) did a comprehensive analysis of the 9 years of SG data at Membach, Belgium (MB), that included both local hydrologic observations and a regional water storage model. MB is one of the stations with significant topography above the SG that requires a detailed integration of effects surrounding the gravimeter to understand changes in very local water storage. Due to the mixed hydrology, the net

annual gravity variations of a few μGal are much less than the more typical 4-8 μGal of surface stations. Rainfall was very carefully modeled as it had been found that transient mass attraction associated with local rainfall can produce gravity variations without surface pressure changes, and this feature was incorporated into the MB analysis in a subsequent paper (Meurers et al., 2007).

A very similar situation was found by Imanishi et al. (2006) at the SG station under Mt. Matsushiro, with again gravity variations of up to 5 μGal following rainfall. At another underground SG station, ST just outside Strasbourg in the Rhine Valley, Longuevergne et al. (2009) modeled the near-surface layer above the gravimeter as a soil moisture layer that retained high rainfall from late winter. This also gave a gravity decrease instead of the expected increase seen at a surface station, the annual amplitude being at the 3–4 μGal level.

A rather different situation is found at Wettzell, a surface station in southern Germany, showing a large annual variation of about 10 μGal amplitude that correlates well with hydrologic models (Creutzfeldt et al., 2010a). This site is a fundamental geodetic reference station of worldwide importance with a number of space geodetic instruments and several generations of SG monitoring. It is the only site to date where a lysimeter (a large 1 m diameter, 1.5 m high cylinder) has been installed to measure precipitation, actual ET, soil moisture variation, and deep drainage (Creutzfeldt et al., 2010c). A comparison of the gravity residuals and the total water storage content from the lysimeter and hydrologic data are shown in **Figure 30**. The agreement is all the more remarkable because the series are not fitted using a parametric model (the usual procedure), but the SG data and the gravity-derived series from the hydrologic data are completely independent. Based on these results, in the subsequent work, Creutzfeldt et al. (2012b, 2013, 2014) used SG residuals to estimate water storage changes over a 10-year period to study water storage dynamics in response to climate variability and extremes, as well as to derive storage–baseflow relationships at different catchment scales.

In addition to the gravity changes following individual rain events, seasonal changes in gravity could also be detected in SGs (Boy and Hinderer, 2006), thanks to their low drift; these are in close agreement with global hydrology models for soil moisture and snow such as LadWorld (Milly and Shmakin, 2002) or GLDAS (Rodell et al., 2004) as shown in **Figure 31**. For Cantley, there is a strong snow contribution in winter adding to the soil humidity, while in Wettzell, it is much smaller. For both stations, there is a good agreement between the gravity residuals and the estimated continental water storage loading effects. In fact, for more than half of the 20 analyzed SGs, there is such a good correlation, as shown in **Figure 32**. For the other stations, the discrepancies may be associated with local hydrology effects, especially when a station is partly underground like Moxa or Strasbourg; indeed, for these stations, there are additional complications (Kroner et al., 2007) due to the gravity effect of water masses above the sensor.

This hydrologic seasonal signal in Europe of a few μGal according to the stations was the basis of recent studies that attempt to correlate ground and satellite gravity data (Crossley et al., 2012; Neumeyer et al., 2006; Weise et al., 2012). Even some variability in the seasonal component is visible in gravity; for instance, the drought impact in Europe during the summer of 2003 was well captured in some European SGs (Andersen and Hinderer, 2005; Hinderer et al., 2006).

Let us also mention the series of papers (Richter et al., 2004; Romagnoli et al., 2003; Zerbini et al., 2001, 2002) that carefully examined both height and gravity changes from continuous GPS and SG (C023) observations at the station Medicina (Italy). These studies, from which we show an example in **Figure 33**, provide a convincing interpretation of the seasonal signal from the combined loading contributions of air pressure, ocean circulation, and hydrology (surficial water table) illustrating once more the need to have simultaneous height and gravity measurements in addition to the monitoring of various environmental parameters in the close vicinity of the station. A similar study combining the analysis of vertical GPS and SG was done for the Strasbourg station by Rosat et al. (2009b). More insight on the variability of the ratio of gravity

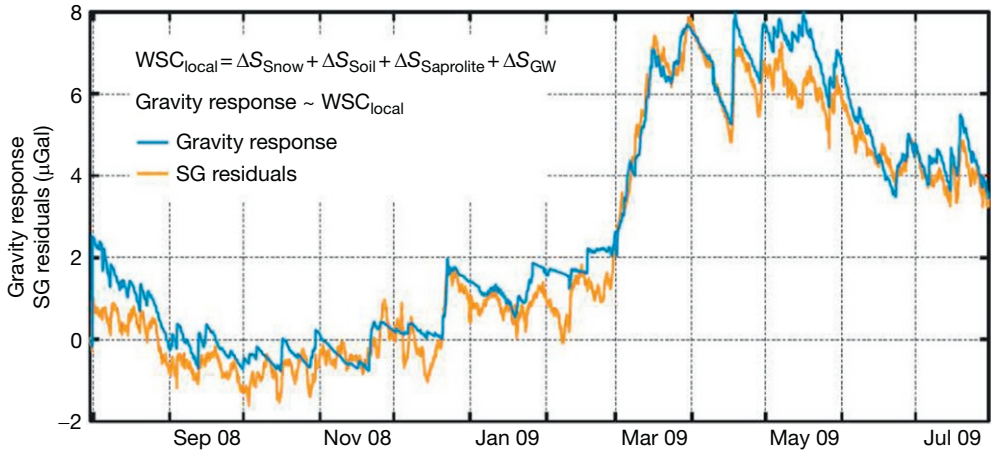


Figure 30 Time variations of the SG residuals and total hydrologic gravity response derived with the lysimeter approach. Reproduced from Creutzfeldt B, Güntner A, Wziontek H, and Merz B (2010) Reducing local hydrology from high precision gravity measurements: A lysimeter-based approach. *Geophysical Journal International* 183(1):178–187, http://dx.doi.org/10.1111/j.1365-246X.2010.04742.x.

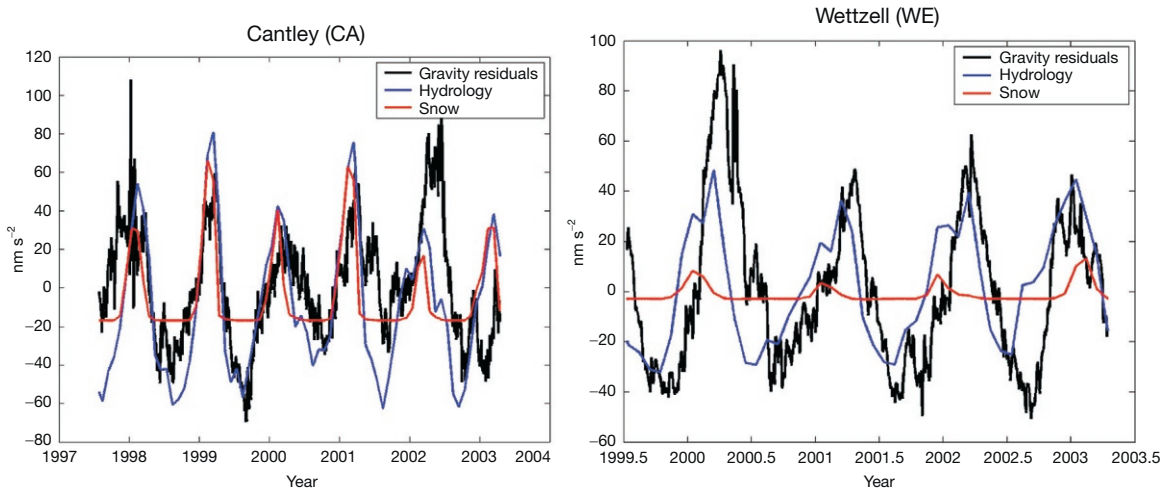


Figure 31 Gravity residuals, hydrology (soil moisture + snow), and snow-modeled contributions at Cantley (Canada) and Wettzell (Germany) stations. Reproduced from Boy J-P and Hinderer J (2006) Study of the seasonal gravity signal in superconducting gravimeter data. *Journal of Geodynamics* 41: 227–233.

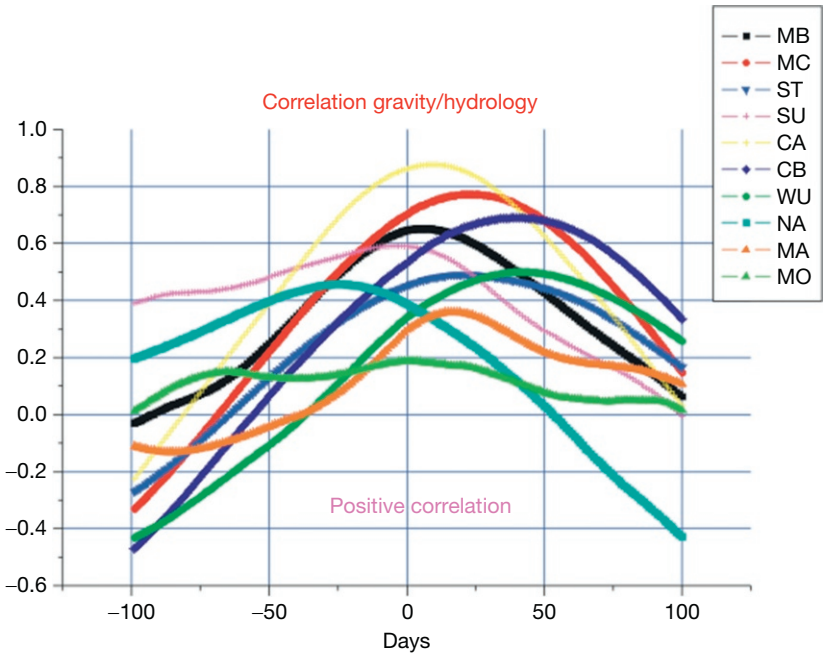


Figure 32 Correlation between gravity and hydrology (continental soil moisture + snow) for some GGP stations; there is a clear positive value with a time lag depending on the station location.

changes and vertical displacement due to surface loads including hydrology can be found in de Linage et al. (2007, 2009a).

Although SGs are the best choice to detect gravity changes at interannual and seasonal timescales because of their high precision and low drift, they only deliver point measurements. This is why SGs were operated in the last years more and more in close association with AGs and RGs (relative spring gravimeters) to provide a hybrid approach to hydrogravimetry.

The advantage of spring gravimeters is they can measure with a high spatial density. In microgravimetric studies, the disadvantage is their large (and unstable) drift. To overcome this problem, one has to frequently come back to a fixed reference point (a typical loop of a few hours). This method is hence expensive in terms of time and manpower. The optimal solution is of course to use a fixed SG as reference point since one knows then exactly the gravity changes at that point without relying on a specific model for tides and air pressure for instance. After drift removal, the time evolution of gravity can then be compared between repetition stations and leads to a spatial view of these changes. The hybrid approach (SG, AG, and RG) is promising as shown by several examples in the succeeding text.

A first example of what can be done surrounding an SG is the work done at Moxa Observatory (MO) in central Germany

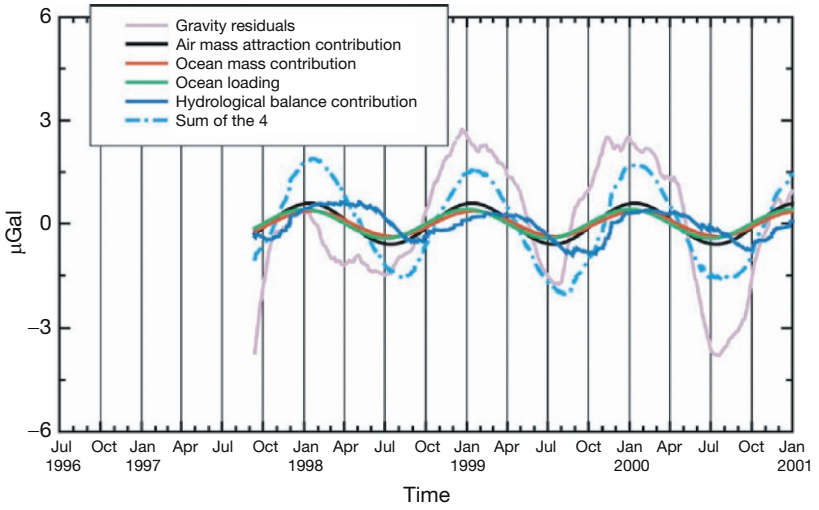


Figure 33 Observed and modeled seasonal gravity changes at Medicina, Italy. Reproduced from Zerbini S, Negusini M, Romagnoli C, Domenichini F, Richter B, and Simon D (2002) Multi-parameter continuous observations to detect ground deformation and to study environmental variability impacts. *Global and Planetary Changes* 34: 37–58.

using a combination of an SG, AG, and several spring meters. The studies were started by Kroner (2001) as a traditional analysis of groundwater variations at an SG site and continued by Kroner and Jahr (2006) who realized an active experiment, including an injection where a large quantity of water (using local fire department equipment) was sprayed onto the vegetated roof of the SG building to saturate the ground and recording the gravity effect with the SG. Clear signals of several μGal were observed, which subsequently led to a decision to remove the vegetation and soil above the instrument. Nevertheless, the SG at MO sits at the bottom of a hillslope that cannot so easily be modified, and the complex nature of the hydrology suggested the more intensive investigation reported by Naujoks et al. (2010). In a major effort over 3 years, up to 400 stations were occupied around the SG and detailed modeling enabled a variety of hydrologic model parameters to be constrained by the gravity observations. Studies continue at MO going a step further to propose consistent hydrologic modeling constrained by gravity (Hasan et al., 2006, 2008).

A second example comes from the Larzac region near Montpelier, France, where time-lapse gravity measurements made with both AGs and Scintrex CG-5s are used to study the hydrology of the Durzon karst (porous sedimentary rock). Gravity variations up to 10 μGal are clearly related to rainfall and the induced infiltration in the karst (Jacob et al., 2008). In June 2011, *i*SG 002 was installed to continuously monitor the gravity changes more precisely and with much better time resolution. In addition, the presence of subsurface caves allowed Jacob et al. (2009) to explore the third (depth) dimension using RGs.

A third example comes from hydrology studies in Africa. The purpose of the GHYRAF (Gravity and Hydrology in Africa) project is to relate gravity and hydrology in West Africa, from an N–S transect that extends from Tamanrasset in the Sahara desert to Djougou (Benin) in the monsoonal region with heavy rains (more than 1.2 m yearly) (Hinderer et al., 2012). The work began by first using FG5 and A10 gravimeters as reference, with additional Scintrex CG-5 gravimeters on nearby repetition networks. There is a fair agreement in the investigated Niger site (Wankama, north of Niamey) in the Sahelian belt as well at in the Benin site (Djougou) between observed FG5 variations (caused by rainfall) and modeled gravity changes using various available hydrologic measurements (piezometers, soil moisture probes, and neutron probes) (Hinderer et al., 2012; Pfeffer et al., 2011). Once again, these promising results led to install an SG in Djougou in order to increase the gravity monitoring of the hydrologic processes. Let us also mention the work done with *in situ* microgravity measurements in the Okavango Delta in Botswana (Christiansen et al., 2011a). The use of time-lapse gravity measurements to calibrate hydrologic model parameters such as specific yield has been positively evaluated by Christiansen et al. (2011b).

A final example is the large-scale test of the gravity effect caused by the controlled water flow at the Southern Avra Valley Storage and Recovery Project (SAVSARP) project in Tucson, Arizona, the United States, at a municipal facility recharging 75 000 000 m^3 annually. The goal is to improve the knowledge of the total subsurface water flow and hydrologic parameters that cannot be identified from hydrology measurements alone. Water is stored and released in large basins on a controlled schedule, and the integrated gravity effect is monitored with two *i*SGs, an A10 and RGs (Creutzfeldt et al., 2012a). In **Figure 34**, the orange lines show residual gravity from *i*SG 004 operating next to recharge pond RB-207 and *i*SG 006 operating 500 m to the south next to pond RB-207. In both *i*SGs, drift-free operation is verified by extensive AG measurements shown as orange points with error bars. The blue line shows ground water level for wells SA-011A and SA-015A, and the blue points with error bars are AG measurements at the water wells SA-011A and SA-015A. These data show how gravity variations can vary dramatically at an active site over short spatial distances. A notable result of this experiment is a demonstration of the usefulness of paired gravimeters to isolate the region of sensitivity – if the signals from two gravimeters in close proximity are differenced, phenomena that affect both

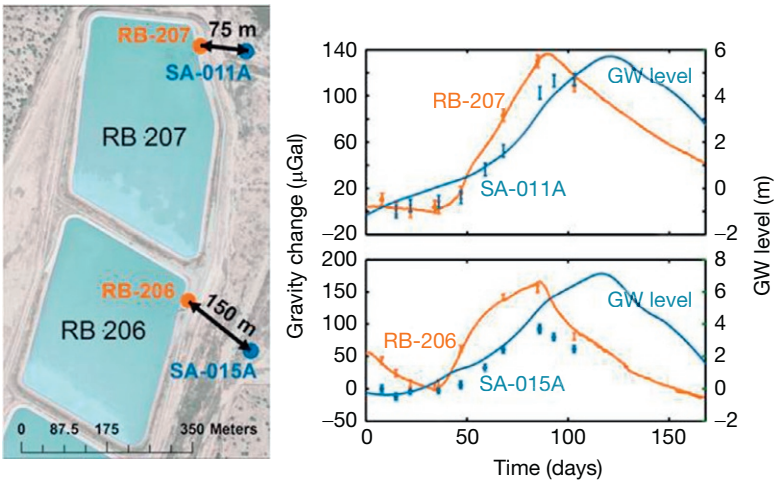


Figure 34 Comparison of SG, AG, and groundwater at the SAVSARP project in Arizona.

equally are canceled out, and the remaining signal is influenced only by nearby mass changes. Furthermore, by changing the spacing and geometric configuration of the gravimeters, the region of sensitivity can be 'focused' to the area of hydrologic interest (Kennedy et al., 2014) .

All these examples show that hybrid (SG, AG, and RG) time-lapse gravity is a useful tool to assess water storage changes and brings more integrative information on the underground water redistribution.

3.04.3.7 Ice Melting

The detection of low-frequency gravity changes due to ice melting is not easy because many different processes act simultaneously and lead to gravity changes that appear as trends: hydrology, tectonics, glacial isostatic adjustment (GIA), and present-day ice melting (PDIM).

To follow the formation and loss of ice sheets and glaciers mainly in polar or high-latitude regions (Antarctica, Arctic, Fennoscandia, and Laurentides), mostly repeated AG measurements are a valuable source of information, especially when some mass transfer occurs (Mazzotti et al., 2007; Sato et al., 2006b). When observing these gravity changes simultaneously with vertical displacement changes (using GPS or VLBI observations), one can infer a value for the admittance of gravity to height changes dg/dh (in μGal cm^{-1}), which can then be interpreted in terms of geodynamic processes.

On long timescales (a few tens to thousands of years), the behavior of the mantle becomes viscoelastic and many studies have shown that, for a large variety of models, viscoelastic loading leads to a ratio of $dg/dh = -1.5$ μGal cm^{-1} (Wahr et al., 1995), which is almost half the value of the elastic loading ratio close to -2.6 μGal cm^{-1} (de Linage et al., 2007; James and Ivins, 1998). In fact, in cases where both elastic and viscoelastic effects act together due to a combination of PDIM and GIA, combining gravity and height change measurements is very efficient in discriminating PGR effects from the present-day ice melting contributions as shown by a recent investigation of the long-term gravity changes in Svalbard (Sato et al., 2006a,b).

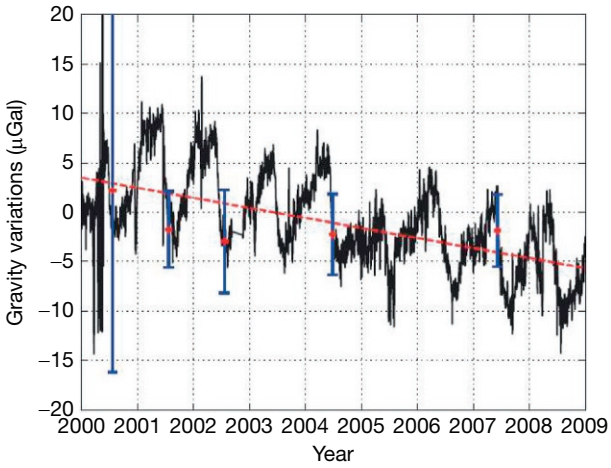


Figure 35 Superposition of relative gravity observations with an SG (black curve) and AG measurements (red cross) at Ny-Ålesund, Svalbard. The red dashed line is the AG linear trend. Reproduced from Mémin A, Rogister Y, Hinderer J, Omang OC, and Luck B (2011) Secular gravity variation at Svalbard (Norway) from ground observations and GRACE satellite data. *Geophysical Journal International* 184: 1119–1130.

More recently, Mémin et al. (2011, 2012) cautioned that in some circumstances, the ratio can be more difficult to interpret unambiguously unless certain approximations are made. Figure 35 shows the seasonal and long-term component of surface gravity as seen by AG episodic measurements and a permanent SG at the Ny-Ålesund station in Svalbard. It is noticeable that the AG points superimpose nicely on the continuous curve provided by the SG, the variability of the annual term explaining why the AG points do not perfectly align on the linear trend. In fact, the geophysical trend is not constant but may vary in time according to a variable melting rate as shown by Omang and Kierulf (2011). A recent paper by Van Camp et al. (2013) explores the optimum method of combining episodic AG and continuous SG data to assess secular gravity changes due to ice melting or tectonics.

3.04.3.8 Earth Rotation

At the long-period end of the observable spectrum of gravity (**Figure 18**), there are two isolated theoretical periods of the Earth's normal-mode spectrum. The first is the ICW that can be computed using either the angular momentum approach or the normal-mode wobble equations. With the former method, Mathews et al. (2002) quoted a value of 6.6 years that is derived from their theory in conjunction with VLBI observations and models, but at this long period, the predicted small amplitude makes this a difficult target (e.g., Guo et al., 2006). A more recent estimate using a Lagrangian formulation of the wobble equations and including elasticity of the IC yields a period closer to 7.5 years for PREM (Rochester and Crossley, 2009).

The other mode is much more accessible and is of course the CW, with a period of about 435 days in the mantle reference frame. This is one of the two components of polar motion normally seen in gravity studies, the other being the smaller amplitude forced annual wobble that is seen in combination with the annual tide and other seasonal effects. Note that at much longer periods, there is an 18.6 year lunisolar tide that will be extremely difficult to identify in gravity because of the needed length of continuous record and the increasing noise with period. Discussion of Richter's (1983) observation of the CW has been mentioned earlier, and almost every SG in the GGP network has reported a clear signal of the CW – see, for example, Richter et al. (1995b) and Sato et al. (1997b) for the Japanese Antarctic station. Here, we review only a few results.

Loyer et al. (1999) showed the importance of using a long data set when trying to infer the transfer function of the polar motion; clearly, more than 6.5 years of continuous data are required to separate the annual component (whatever its origin) from the 435-day term. They found a gravimetric amplitude factor $\delta = 1.18 \pm 0.10$ and a phase delay of a few degrees using only the Strasbourg station. Harnisch and Harnisch (2006b) generalized this study by using data from 12 GGP stations with lengths varying between 4 and 18 years to investigate the polar motion contribution to gravity. They found that in general the gravimetric amplitude factors for the Chandler wobble are close to the nominal value of 1.16 and phase lags of a few degrees. Xu et al. (2004b) also made a similar study on five SGs from the GGP network, and we show in **Figure 36** the typically good agreement between gravity residuals and polar motion for the stations Membach and Potsdam. Combining the results leads to a δ factor of 1.16 ± 0.07 and to a phase delay of $-1.30° \pm 1.33°$. Finally, a data set of nine SGs was analyzed for the polar motion response by Ducarme et al. (2006) who found a mean δ factor of 1.179 ± 0.004, similar to Loyer et al. (1999).

A value slightly larger than the nominal value of 1.16 is to be expected when including mantle inelasticity and/or ocean pole tide contribution. The variability on the phase delays between different stations is, however, still large and unexplainable in terms of Earth's rheological properties.

3.04.3.9 Tectonic Effects

Another class of long-term gravity changes is due to active tectonics. Until recently, most studies have been done using AGs (e.g., Niebauer et al., 1995), rather than SGs. One reason for this is that AGs are drift-free and can be physically moved from site to site over a large area with relative ease. On the other hand, SGs incur offsets when moved, and offsets compromise repeated spatial surveys. Nevertheless, it is now recognized that SGs are ideal for studying PGR at a single site when complemented by AG measurements, especially to record unmodeled effects such as hydrology that can affect AG values.

A second reason is that the SGs have a small instrument drift that must be accounted for as part of any estimate of the secular gravity trend. The potential of long-term gravity measurements to solve tectonic problems is however significant, as clearly demonstrated by a study using FG5 AGs on the Cascadia subduction zone (Lambert et al., 2006). From a practical point of view, gravimeters are useful to constrain height changes in general, especially when combined with vertical displacement (GPS and VLBI), that help understand the physical process generating mass redistribution. This is especially true in tectonic areas where there is a significant vertical motion.

The first study showing coseismic effects detected by SGs is the study by Imanishi et al. (2004) of coseismic displacements that accompanied a large $M_w = 8.0$ earthquake in Japan. Using three SGs in a linear array (**Figure 37**), they showed that gravity offsets occurred that matched the theoretical prediction of static displacements that accompany a fault rupture process. The offsets were 0.58, 0.10, and 0.07 μGal at stations (ES, MA, and KY), respectively, in line with the epicenter at Tokachi-Oki (near Hokkaido). Such small offsets (especially close to 0.1 μGal) would hardly be identified in an individual record, but this is the first time a gravity array has performed at this

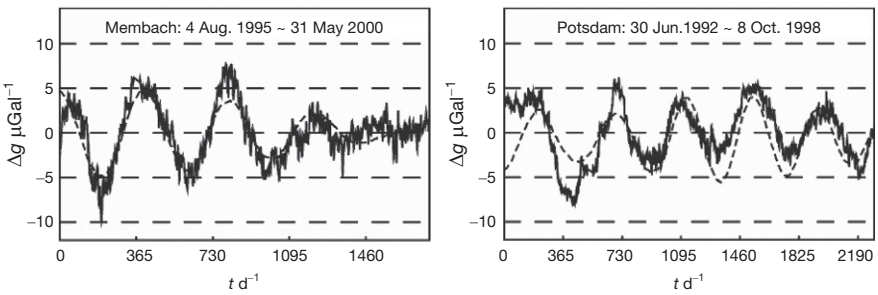


Figure 36 Comparison between gravity residuals (solid line) and polar motion gravity prediction (dashed line) at Membach (Belgium) and Potsdam (Germany) stations. Reproduced from Xu J-Q, Sun H-P, and Yang X-F (2004) A study of gravity variations caused by polar motion using superconducting gravimeter data from the GGP network. *Journal of Geodesy* 78: 201–209, http://dx.doi.org/10.1007/s00190-004-0386-1.

level of precision. This difficulty was indeed revealed by a later attempt with a smaller earthquake (Imanishi et al., 2007). It is hence more probable that gravity studies of coseismic and postseismic displacements rely more on AGs (Tanaka et al., 2001) or satellite gravimetry with GRACE data (de Linage et al., 2009b; Matsuo and Heki, 2011; Panet et al., 2007). In fact, the question of monitoring gravity changes associated with tectonic processes (due to elastic stress accumulation in locked areas and stress release by earthquakes) from GRACE data has been raised quickly after the GRACE launch in 2002 (Mikhailov et al., 2004), but of course, a combination of satellite data with surface SG measurements would give a much stronger data set.

The significance of the coseismic detection by Imanishi et al. (2004) can be transferred to another context, that of subduction-induced silent earthquakes and episodic tremor and slip (ETS) (Rogers and Dragert, 2003) that have been identified off the coast of Japan and also with the Cascadia subduction zone (Dragert et al., 2001, 2004). Episodic tremor that accompanies the silent slip events can last for several days, but there is no identifiable earthquake. Such events are identified primarily as horizontal displacements on GPS arrays, but Lambert et al. (2006) had shown that AG measurements also reveal coincident offsets in gravity. The significant reduction of error for linear gravity trend estimates from repeated AG measurements at Ucluelet (**Figure 38**) when repeated ETS 'steps' are allowed for suggested that slow-slip events beneath Vancouver Island have an associated gravity signal of several μGals. Consequently, in July 2012, *i*SG 001 was installed in the seismic vault at the Pacific Geoscience Centre located on southern Vancouver Island near Sidney, BC. The objective is to clearly resolve gravity signals associated with ETS events that reoccur roughly every 15 months in this region and determine if local mass changes occur.

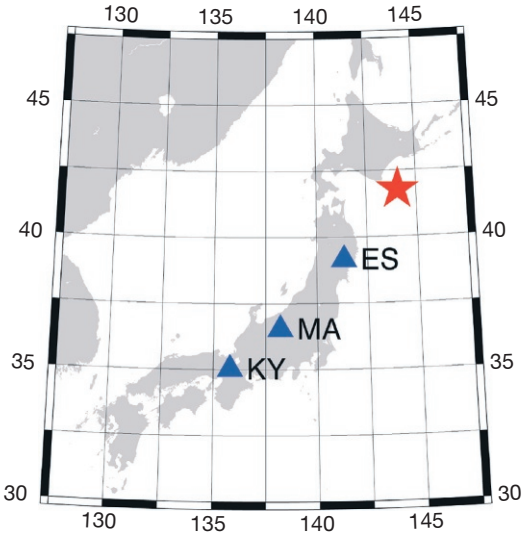


Figure 37 Japan SG array (blue stations) that observed gravity offsets due to a local large earthquake (red).

Another study based on 7 years of collocated gravity measurements of SG and AG at Membach (Belgium) by Francis et al. (2004) indicates that there is a small decrease in gravity connected to uplift seen by GPS at the same location. A longer data set is required to determine if this gravity decrease is due to PGR or active tectonics in the Ardennes mountains (see also Van Camp et al., 2002).

Such comparisons of AG and GPS measurements are much easier to do when there are continuous SG measurements at the same location. As Van Camp et al. (2013) showed, the continuity of the SG record is often important to check the integrity of the AG measurements (discussed earlier in connection with the SG calibration) and to model effects such as hydrology that increase the AG scatter. Amalvict et al. (2004, 2006) analyzed the long-term gravity changes of the Strasbourg SG with regular AG measurements and collocated GPS and interpreted the results in terms of hydrology and tectonics of the Rhine Graben. **Figure 39** shows the gravity trend that seems to be present in an 8-year data set (1997–2004) and how hydrologic contributions partly explain some of the gravity features. An updated comparison of the AG and SG observations at Strasbourg can be found in Rosat et al. (2009b).

Mouyen et al. (2009) reported on a project to determine the uplift of south central Taiwan from a combination of AG and

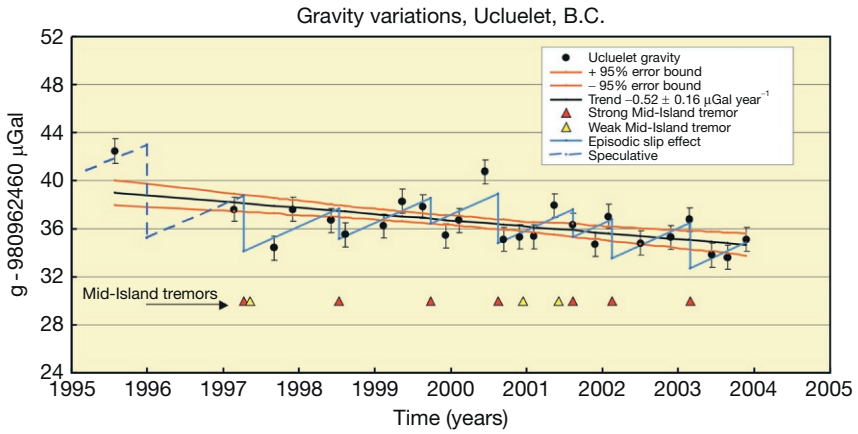


Figure 38 Variations in AG at Ucluelet (western coast of Vancouver Island) showing some concordance with the episodic slip and seismic tremor activity above the Cascadia subduction zone. Courtesy of T. Lambert.

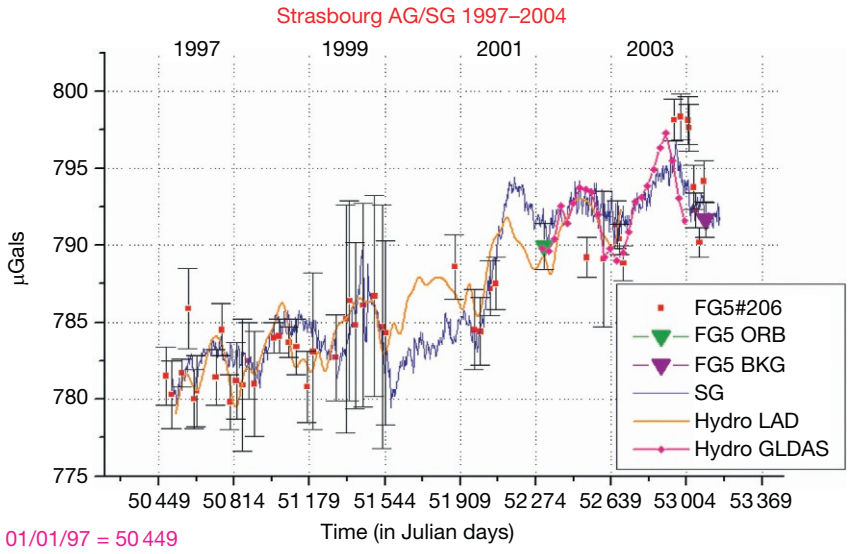


Figure 39 An example of SG/AG superposition at Strasbourg (France) from 1997 to 2004. The gravity observations are corrected for tides (solid + ocean load), air pressure, and polar motion; the instrumental drift of the SG has been removed and hydrologic contributions added. Reproduced from Amalvict M, Hinderer J, and Rozsa S (2006) Crustal vertical motion along a profile crossing the Rhine graben from the Vosges to the Black Forest Mountains: Results from absolute gravity, GPS and levelling observations. *Journal of Geodynamics* 41: 358–368.

GPS measurements. With ten AG stations now surveyed annually over a period of 5 years, the uplift at some stations is as large as 12 mm year^{-1}, but the expected gravity decrease is not seen at all stations, due in this case to competing local effects such as landslides and erosion (Mouyen et al., 2012).

3.04.3.10 Ground/Satellite Gravity Field Comparison

There has been a major international effort in the past two decades to significantly improve the resolution of the Earth's global gravity field using low-altitude satellites. The first satellite CHAMP was launched in 2000 to improve the static field and was followed 2 years later by GRACE that was designed for time-varying gravity. The third mission, GOCE (Gravity Field and Steady-State Ocean Circulation Explorer), was launched in 2009 to map the static field at high resolution using a three-axis gravity gradiometer. Both CHAMP and GRACE had orbits at about 400 km altitude, but the GOCE orbit is at 260 km altitude, much closer to the ground than any previous mission and hence more sensitive to small-scale gravity changes. GOCE has already had significant success in improving the knowledge of the static field to a resolution of about 100 km on the ground (e.g., Hirt et al., 2012).

The primary goal of the GRACE mission, which continues well past its envisaged lifetime, is to use the temporal changes of the Earth's gravity field to infer changes in regional and continental water storage and ocean circulation (see Rodell and Famiglietti, 1999; Wahr et al., 1998).

A major concern of all satellite measurements is how to calibrate and validate gravity data using independent observations or models. For time-variable gravity, in addition to comparisons with models (the primary technique used to date), there are several possibilities using actual measurements at the ground (GPS, gravity) or at the ocean bottom (water pressure). This problem is important because if successful, it would produce an independent method of validation that does not rely solely on modeling. We will show hereafter that surface measurements from the GGP network provide a useful additional constraint on space gravity data. The validation signal is related to seasonal gravity changes that are coherent on length scales appropriate to satellite altitudes (typically a few hundreds of km).

A first study directly comparing CHAMP data to six SG ground observations was done by Neumeyer et al. (2004b) and has led to satisfactory results for all the stations in the 1-year analysis period (from December 2000 to December 2001). The superposition of the monthly gravity mean values from the SG residuals (after correction for solid tides, ocean and atmospheric loadings, and polar motion) with the CHAMP reconstructed values at the SG sites is rather good. Neumeyer et al. (2006) further extended this study to GRACE data pointing out again the partial agreement between surface gravity and satellite-derived gravity at specific locations. Before a detailed comparison can be made, however, one has to remember that ground gravity measurements include necessarily a contribution from the vertical motion of the instrument through the ambient gravity field. This signal does not affect the orbiting satellite, and hence, there is a difference in the gravity changes as seen at (moving) ground level and by the satellite (Hinderer et al., 2006).

We note in these studies that the comparison of single-station results with the large-scale satellite solutions is problematic due to the completely different error budgets involved. GRACE data, for example, are good to 1 µGal only over length scales longer than 500–1000 km, whereas SGs are good to the same accuracy (or better) at a single point. In order to average SG measurements and reduce local effects, there have been attempts to assemble a network solution from nearby SG stations rather than doing the aforementioned single-station comparison. Within the existing rather sparse GGP network,

Europe is obviously the best place to try such an approach (Figure 17).

The approach was first initiated using 1 year of SG data by Crossley and Hinderer (2002) and Crossley et al. (2003) and extended to longer data sets by Crossley et al. (2003, 2004, 2005). This approach was further extended to a 21-month time interval to intercompare surface data (GGP European subnetwork), satellite data from GRACE, and theoretical predictions for two hydrology models LAD (Milly and Shmakin, 2002) and GLDAS (Rodell et al., 2004) (Andersen et al., 2005a; Hinderer et al., 2006). The results show the existence of an annual signal that is coherent over Europe with an amplitude of a few μGal mostly due to the seasonal loading from continental hydrology (soil moisture + snow) according to recent models such as LAD or GLDAS. There is even a possibility to detect in GRACE data interannual signals (Andersen and Hinderer, 2005) and, in particular, there is a clear evidence that GRACE has been affected by the heat wave that occurred in summer 2003 in Europe (Andersen et al., 2005b). The Wettzell (Germany) and the Medicina (Italy) SG data seem to confirm this point as shown by Figure 40.

The comparison of ground–satellite gravity for Europe has subsequently been revisited by several groups. Neumeyer et al. (2008) made corrections for local hydrology at the stations and analyzed the residuals using station EOFs, with good results with a strong common annual signal. Weise et al. (2009), in a similar study, did not use the EOF technique, preferring to compare solutions at each station. There exist some interesting differences of opinion as to what component of local hydrology should be corrected from the ground stations when making such comparisons. Crossley et al. (2012) did not correct for local hydrology, but still found a 79% agreement (Figure 41) between the first EOF mode from seven GGP stations (1σ error of 1.0 μGal) and the improved GRACE satellite

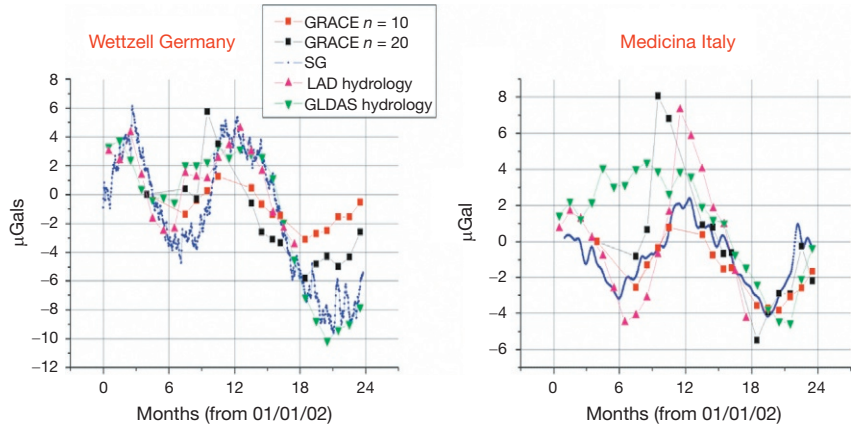


Figure 40 An example of the superposition of gravity changes seen by GRACE (in red for $n = 10$ and in black for $n = 20$) and the superconducting gravimeter in Wettzell (Germany) and Medicina (Italy) (in blue) and contributions from continental hydrology models like GLDAS (in green) and LADWorld (in purple).

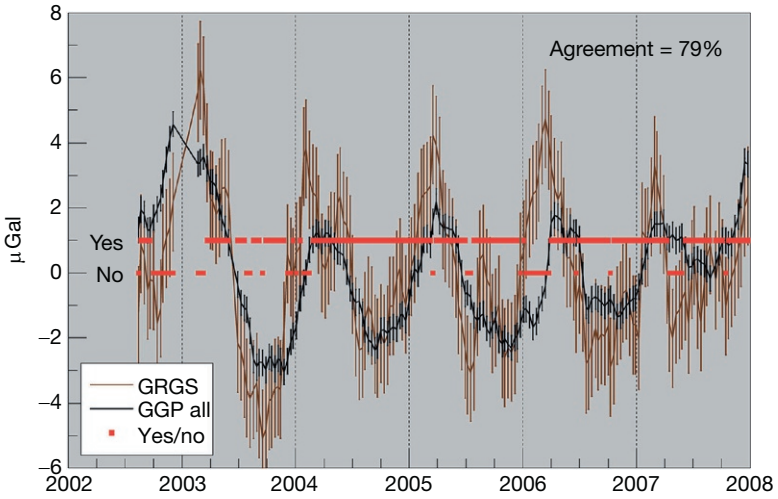


Figure 41 Quantitative agreement between ground stations (the seven GGP stations in Europe) and GRACE satellite data (GRGS solutions) during the 2002–08 time period. The yes/no bars stand for the superposition (or lack of) of the observations within respective error bars (1σ error of 1 μGal for GGP and 1.5 μGal for GRGS). Reproduced from Crossley D, de Linage C, Boy J-P, Hinderer J, and Famiglietti J (2012) A comparison of the gravity field over Central Europe from superconducting gravimeters, GRACE and global hydrological models, using EOF analysis. *Geophysical Journal International* 189: 877–897.

solutions (1σ error of 1.5 μGal) from the GRGS/CNES group (Bruinsma et al., 2010). If only the three GGP aboveground stations are used, this agreement reaches 90%.

Recently, two rather similar papers by Weise et al. (2012) and Abe et al. (2012) more or less came to the same conclusions, and so all studies agree on the strong common seasonal variations that are seen at the same level by the ground and GRACE data. These generally conform, with some small disagreement about amplitudes, with global hydrologic models such as GLDAS and WGHM (Water Gap Hydrological Model; Döll et al., 2003).

3.04.3.11 Future Possibilities

The development of the *i*SG points to several more 'practical' applications of SGs that previously have been served almost exclusively by spring gravity meters, both portable and tidal, and AGs. These include fundamental physics and metrology, glaciology, hydrology, 4-D continuous monitoring of CO_2 sequestration and enhanced oil reservoirs, monitoring volcanoes, and geothermal applications. Design and construction of a shallow borehole SG are under way and discussions are taking place to design and operate SGs on the ocean floor.

In February 2013, installation of *i*SG 005 started in the Laboratoire National de Métrologie et d'Essais in Trappes, France, for studies of the long-term stability of the absolute cold atom gravimeter (CAG) (Bodart et al., 2010). This CAG, which participates in the International Comparison of Absolute Gravimeters, is part of the French watt balance project aiming to redefine the kg unit (Merlet et al., 2008). In particular, this SG will provide a continuous measurement of *g* in the watt balance laboratory. Inquiries from institutions in Canada, China, and the United States indicate that SGs will soon be used in many worldwide metrology laboratories.

In June 2012, *i*SG 003 was installed in Juneau, Alaska, as a part of a collaborative project among Universities of Tohoku, Tokyo (Japan), and Alaska (the United States), called ISEA (International Geodetic Project in SE Alaska) since 2005. They added the SG measurements to the previous absolute gravity, GRACE, and GPS data to measure GIA and to estimate the viscosity structure of the Earth's mantle (Sato et al., 2012; Sun et al., 2010). Also of recent interest is the proposal to use AG measurements in conjunction with GRACE data to monitor Fennoscandian uplift, and such a project would benefit considerably from one or two stations that have an SG for continuous monitoring (Timmen et al., 2012).

In December 2012, *i*SG 010 was installed in the Farnsworth field, Texas, to take part in a joint Japan and US project to study CO_2 sequestration and enhanced oil recovery. Three kinds of gravimetric baseline measurements are being made: continuous gravity measurements with the *i*SG, colocated AG measurements, and time-lapse gravity over the Farnsworth field with RGs.

Hydrology will continue to evolve as one of the important targets in geodetic and geophysical projects, with *i*SGs used for surface measurements, while a borehole SG instrument is being developed for subsurface hydrologic applications. The borehole dewar design is 27 cm in diameter, holds about 2 l of liquid; and will operate in a 30 cm outer diameter casing up to a depth of 50 m. At this depth, it is most practical to operate the compressor at the surface and route compressor hoses down the hole. Nonetheless, it is also possible – and necessary at greater depths – to reconfigure the compressor to fit inside the casing and to lower it in tandem with the sensor. In initial experiments, two SGs will operate continuously – one on the surface and one positioned below the subsurface hydrology – and the difference signal will be used to isolate the near-surface water storage.

By combining SG surface and borehole gravimeters, it would be possible to monitor mass change in discrete intervals throughout the subsurface, including separating land-surface hydrology from deep aquifer recharge, identifying displacement of water by injected CO_2 in carbon sequestration projects, and monitoring dewatering during aquifer pumping. This geometry would also solve the problem of transferring the SG signal from any underground location (e.g., J9 in Strasbourg) to a ground surface point for the comparison of ground gravity with GRACE, as proposed at the end of the paper by Crossley et al. (2012).

Several groups are considering geothermal monitoring, which is a logical extension of the hydrologic applications. In fact, more than 30 years ago, an SG was used in this context at the Geysers field (Olson and Warburton, 1979) to detect the gravity decrease due to steam production. Time-lapse gravimetry aims to monitor not only the natural behavior of the geothermal reservoir but also the transient stimulation by injection or even the long-term evolution of the reservoir. Several studies introduce the concept of hybrid gravimetry (Oka et al., 2012; Sugihara and Ishido, 2008; Sugihara and Nawa, 2012; Sugihara et al., 2013), which rely on the combined use of different types of gravimeters (AG and RGs) to monitor geothermal reservoirs. The addition of SG measurements will provide continuous monitoring of the time-varying gravity in a precise and stable way, while AG episodic measurements keep the absolute gravity reference and the RGs bring additional insight on spatial gravity changes by repetition network observations (e.g., Crossley and Hinderer, 2005).

SGs will soon be deployed for monitoring volcanoes – both their slow deformation and their explosive activity. Continuous high-quality gravity measurements at volcanoes – in combination with geodetic ground deformation and seismic observations – may enable detection of subsurface mass redistribution long before other eruption precursors appear (Carbone and Poland, 2012; Carbone et al., 2006, 2008, 2012; Rymer and Locke, 1995). Hence, high-quality gravity measurements are a key tool for long-term hazard assessment of volcanoes as discussed by Riccardi et al. (2008). For this reason, several groups in Europe, Mexico, and South America are interested in *i*SGs for volcano monitoring. This year, it is planned to install an *i*SG in the summit zone of Mt. Etna in the framework of a scientific project managed by INGV; for a comprehensive review of Etna gravity, see Carbone and Greco (2007).

Carbone et al. (2003) reported several months of gravity observations along a NS profile on the flank of Mt. Etna that showed variations of more than 50 μGal using Scintrex portable instruments and a continuously operating LaCoste–Romberg base station. Had the base station been equipped with an SG, the results would have been even more impressive. Using a LaCoste–Romberg gravimeter continuously recording at Kilauea

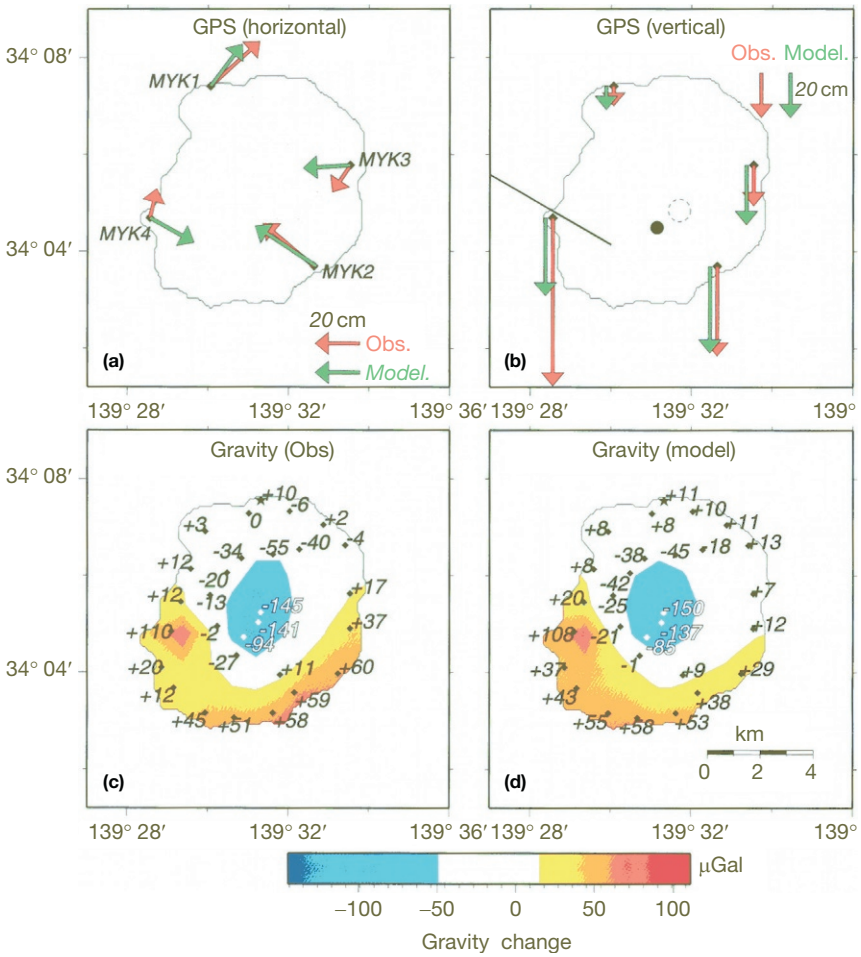


Figure 42 Observed and modeled combinations of GPS and gravity variations, prior to the caldera collapse at Miyake-jima volcano, Japan, in 2000. Reproduced from Furuya M, Okubo S, Sun W, Tanake Y, Oikawa J, and Watanabe H (2003) Spatio-temporal gravity changes at Miyakejima Volcano, Japan: Caldera collapse, explosive eruptions, and magma movement. *Journal of Geophysical Research* 108(B4): 2219.

volcano, Carbone and Poland (2012) succeeded in observing small gravity fluctuations (amplitude of a few μGal and period of a few minutes) that they interpreted as due to shallow magma convection in a volume of known magma storage. An SG would have detected this signal much more easily and revealed finer details of the signal itself.

Both gravity and GPS measurements were accomplished by Furuya et al. (2003) – shown in **Figure 42** – and this allows a more detailed interpretation of the subsurface mass changes than using either method alone. The gravity signals are very large (~100 μGal) compared with those discussed elsewhere in this chapter, and there would be no question of their detection by an SG, even without detailed AG backup measurements.

3.04.3.12 Conclusions

We now summarize the main points of the use of SGs in geodetic measurements:

(1) Over the 40 years since it was first developed, the SG has proved to be an extremely reliable instrument for determining gravity variations from 1 s to periods of several years, and several instruments have been operated for significantly longer than a decade without interruption.
(2) Calibration changes in SGs are virtually nonexistent, and drift rates are at the level of a few μGal per year, making them ideal for long-term monitoring of the gravity field (tectonics, seasonal changes, and polar motion).
(3) The precision of 0.1 nGal in frequency domain and an accuracy of better than 0.1 μGal in the time domain have contributed to major improvements in tidal analysis, ocean loading computations and atmospheric effects that were not possible with other types of gravimeter.
(4) With SGs, it is now possible to seriously model the gravity effects in hydrology due to rainfall and to quantify these effects in soil moisture, groundwater variations, and atmospheric mass changes.
(5) The SG has proved to be even more useful than spring ET meters for the study of the Earth's free oscillations, particularly at periods longer than 500 s.
(6) The combination of SGs and satellite measurements of the gravity field is proving to be consistent and complementary indicators of regional hydrology.

(7) New possibilities exist for the use of SGs in hydrology, geothermal applications, metrology, tectonics, and the monitoring of volcanoes, particularly with the convenient operation of the new *i*SG.
(8) Future studies in gravity would be much enhanced if SGs can be deployed in array form and combined, where appropriate, with AGs, portable gravimeters, and GPS receivers.

Acknowledgments

Much of the development of the SG has been carried out at GWR by contracts initiated by Bernd Richter and funded by the Institut für Angewandte Geodäsie (IFAG, Frankfurt, Germany) and later by the Bundesamt für Kartographie und Geodäsie (BKG, Frankfurt).
Without this involvement, the commercial venture would have been difficult to maintain. We also acknowledge the support of the US Department of Energy under Grant No. DE-FG03-95ER81979 and the US National Science Foundation under Grant No. 9529827. This research was supported through NSF EAR Award #0409381.

References

Abe M, Kroner C, Förste C, et al. (2012) A comparison of GRACE-derived temporal gravity variations with observations of six European superconducting gravimeters. *Geophysical Journal International* 191(2): 545–556. http://dx.doi.org/10.1111/j.1365-246X.2012.05641.x.

Abe M, Kroner C, Neumeyer J, and Chen XD (2010) Assessment of atmospheric reductions for terrestrial gravity observations. *Bulletin d'Informations des Marees Terrestres* 146: 11817–11838.

Abe M, Takemoto S, Fukuda Y, et al. (2006) Hydrological effects on the superconducting gravimeter observation in Bandung. *Journal of Geodynamics* 41(1–3): 288–295. http://dx.doi.org/10.1016/j.jog.2005.08.030.

Achilli V, Baldi P, Casula G, et al. (1995) A calibration system for superconducting gravimeters. *Bulletin Géodésique* 69: 73–80.

Agnew DC (1997) NLOADF: A program for computing ocean-tide loading. *Journal of Geophysical Research* 102: 5109–5110.

Agnew DC (2012) *SPOTL: Some Programs for Ocean-Tide Loading, SIO Technical Report, Scripps Institution of Oceanography*. Available at: http://escholarship.org/uc/sio_techreport.

Agnew DC and Farrell WE (1978) Self-consistent equilibrium ocean tides. *Geophysical Journal of the Royal Astronomical Society* 55: 171–181.

Amalvict M, Hinderer J, Boy J-P, and Gegout P (2001a) A three year comparison between a superconducting gravimeter (GWR C026) and an absolute gravimeter (FG5#206) in Strasbourg (France). *Journal of the Geodetic Society of Japan* 47: 410–416.

Amalvict M, Hinderer J, Gegout P, Rosat S, and Crossley D (2002) On the use of AG data to calibrate SG instruments in the GGP network. *Bulletin d'Informations des Marees Terrestres* 135: 10621–10626.

Amalvict M, Hinderer J, Makinen J, Rosat S, and Rogister Y (2004) Long-term and seasonal gravity changes and their relation to crustal deformation and hydrology. *Journal of Geodynamics* 38: 343–353.

Amalvict M, Hinderer J, and Rozsa S (2006) Crustal vertical motion along a profile crossing the Rhine graben from the Vosges to the Black Forest Mountains: Results from absolute gravity, GPS and levelling observations. *Journal of Geodynamics* 41: 358–368.

Amalvict M, McQueen H, and Govind R (2001b) Absolute gravity measurements and calibration of SG-CT031 at Canberra, 1999–2000. *Journal of the Geodetic Society of Japan* 47: 334–340.

Andersen O and Hinderer J (2005) Global inter-annual gravity changes from GRACE: Early results. *Geophysical Research Letters* 32: L01402. http://dx.doi.org/10.1029/2004GL020948.

Andersen O, Hinderer J, and Lemoine FG (2005a) Seasonal gravity field variations from GRACE and hydrological models. In: Jekeli C, Bastos L, and Fernandes J (eds.) *Gravity, Geoid and Space Missions. IAG Symposia*, vol. 129, pp. 316–321. Heidelberg: Springer Verlag.

Andersen O, Seneviratne S, Hinderer J, and Viterbo P (2005b) GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophysical Research Letters* 32: L18405. http://dx.doi.org/10.1029/2005GL023574.

Baker T and Bos M (2003) Validating Earth and ocean tide models using tidal gravity measurements. *Geophysical Journal International* 152: 468–485.

Banka D and Crossley D (1999) Noise levels of superconducting gravimeters at seismic frequencies. *Geophysical Journal International* 139: 87–97.

Berger J, Davis P, and Ekström G (2004) Ambient Earth noise: A survey of the Global Seismographic Network. *Journal of Geophysical Research* 109: B11307. http://dx.doi.org/10.1029/204JB003408.

Bodart Q, Merlet S, Malossi N, Pereira Dos Santos F, Bouyer P, and Landragin A (2010) A cold atom pyramidal gravimeter with a single laser beam. *Applied Physics Letters* 96(13): 134101, 1–3.

Bos MS and Baker TF (2005) An estimate of the errors in gravity ocean tide loading computations. *Journal of Geodesy* 79: 50–63.

Bos M, Baker TF, Lyard FH, Zürn W, and Rydelek PA (2000) Long-period lunar tides at the geographic South Pole and recent models of ocean tides. *Geophysical Journal International* 143: 490–494.

Bower DR and Courtier N (1998) Precipitation effects on gravity measurements at the Canadian absolute gravity site. *Physics of the Earth and Planetary Interiors* 106: 353–369.

Bower DR, Liard J, Crossley D, and Bastien R (1991) Preliminary calibration and drift assessment of the superconducting gravimeter GWR12 through comparison with the absolute gravimeter JILA 2. In: *Proceedings of the Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Europeen de Geodynamique et de Seismologie, 3, Luxembourg* 129–142.

Boy J-P and Chao BF (2005) Precise evaluation of atmospheric loading effects on Earth's time-variable gravity field. *Journal of Geophysical Research* 110: B08412. http://dx.doi.org/10.1029/2002JB002333.

Boy J-P, Gegout P, and Hinderer J (2001) Gravity variations and global pressure loading. *Journal of the Geodetic Society of Japan* 47(1): 267–272.

Boy J-P, Gegout P, and Hinderer J (2002) Reduction of surface gravity data from global atmospheric pressure loading. *Geophysical Journal International* 149: 534–545.

Boy J-P and Hinderer J (2006) Study of the seasonal gravity signal in superconducting gravimeter data. *Journal of Geodynamics* 41: 227–233.

Boy J-P, Hinderer J, Amalvict M, and Calais E (2000) On the use of long records of superconducting and absolute gravity observations with special application to the Strasbourg station, France. In: *Proceedings of the Workshop: High-Precision Gravity Measurements with Application to Geodynamics and Second GGP Workshop, Cahiers du Centre Européen de Géodynamique et de Séismologie, 17, Luxembourg*, pp. 67–83.

Boy J-P, Hinderer J, and Gegout P (1998) Global atmospheric loading and gravity. *Physics of the Earth and Planetary Interiors* 109: 161–177.

Boy J-P and Liard F (2008) High-frequency non-tidal ocean loading effects on surface gravity measurements. *Geophysical Journal International* 175: 35–45.

Boy J-P, Llubes M, Hinderer J, and Florsch N (2003) A comparison of tidal ocean loading models using superconducting gravimeter data. *Journal of Geophysical Research* 108(B4): 2193. http://dx.doi.org/10.1029/2002JB002050.

Boy J-P, Llubes M, Ray R, Hinderer J, and Florsch N (2006) Validation of long-period oceanic tidal models with superconducting gravimeters. *Journal of Geodynamics* 41: 112–118.

Boy J-P, Llubes M, Ray R, et al. (2004) Non-linear oceanic tides observed by superconducting gravimeters in Europe. *Journal of Geodynamics* 38: 391–405.

Bruinsma S, Lemoine JM, Biancale R, and Vales N (2010) CNES/GRGS 10-day gravity field models (release 2) and their evaluation. *Advances in Space Research* 45: 587–601.

Busse FH (1974) On the free oscillation of the Earth's inner core. *Journal of Geophysical Research* 79: 753–757.

Carbone D and Greco F (2007) Review of microgravity observations at Mt. Etna: A powerful tool to monitor and study active volcanoes. *Pure and Applied Geophysics* 164: 769–790. http://dx.doi.org/10.1007/s00024-007-0194-7, 0033-4553/07/040769-22.

Carbone D, Budetta G, and Greco F (2003) Possible mechanisms of magma redistribution under Mt Etna during the 1994–1999 period detected through microgravity measurements. *Geophysical Journal International* 153: 187–200.

Carbone D and Poland MP (2012) Gravity fluctuations induced by magma convection at Kīlauea Volcano, Hawaii. *Geology* 40: 803–806. http://dx.doi.org/10.1130/G33060.1.

Carbone D, Zuccarello L, Montalto P, and Rymer H (2012) New geophysical insight into the dynamics of Stromboli volcano (Italy). *Gondwana Research* 22: 290–299. http://dx.doi.org/10.1016/j.gr.2011.09.007.

Carbone D, Zuccarello L, and Saccorotti G (2008) Geophysical indications of magma uprising at Mt Etna during the December 2005 to January 2006 non-eruptive period. *Geophysical Research Letters* 35: L06305. http://dx.doi.org/10.1029/2008GL033212.

Carbone D, Zuccarello L, Saccorotti G, and Greco F (2006) Analysis of simultaneous gravity and tremor anomalies observed during the 2002–2003 Etna eruption. *Earth and Planetary Science Letters* 245: 616–629.

Cartwright DE and Edden CA (1973) Corrected tables of spherical harmonics. *Geophysical Journal of the Royal Astronomical Society* 33(3): 253–264.

Cartwright DE and Taylor RJ (1971) New computations of the tide generating potential. *Geophysical Journal of the Royal Astronomical Society* 23(1): 45–74.

Christiansen L, Binning PJ, Andersen O, Rosbjerg D, and Bauer-Gottwein P (2011a) Ground-based time-lapse gravity data used to calibrate a river bank storage model in the Okavango Delta, Botswana. In: *Proceeding of International Conference on Groundwater: Our Source of Security in an Uncertain Future, Pretoria, South Africa: International Association of Hydrogeologists (IAH)*.

Christiansen L, Binning PJ, Rosbjerg D, Andersen OB, and Bauer Gottwein P (2011b) Using time lapse gravity for groundwater model calibration: An application to alluvial aquifer storage. *Water Resources Research* 47: W06503. http://dx.doi.org/10.1029/2010WR009859.

Courtier N, Ducarme B, Goodkind J, et al. (2000) Global superconducting gravimeter observations and the search for the translational modes of the inner core. *Physics of the Earth and Planetary Interiors* 117: 3–20.

Creutzfeldt B, Ferré PA, Troch PA, Merz B, Wziontek H, and Güntner A (2012a) Total water storage dynamics in response to climate variability and extremes – Inference from long-term terrestrial gravity measurement. *Journal of Geophysical Research* 117, D08112.

Creutzfeldt B, Güntner A, Klügel T, and Wziontek H (2008) Simulating the influence of water storage changes on the superconducting gravimeter of the geodetic observatory Wettzell, Germany. *Geophysics* 73: WA95–WA104.

Creutzfeldt B, Güntner A, Thoss H, Merz B, and Wziontek H (2010a) Measuring the effect of local water storage changes on in situ gravity observations: Case study of the Geodetic Observatory Wettzell, Germany. *Water Resources Research* 46: W08531. http://dx.doi.org/10.1029/2009WR008359.

Creutzfeldt B, Güntner A, Vorogushyn S, and Merz B (2010b) The benefits of gravimeter observations for modelling water storage changes at the field scale. *Hydrology and Earth System Sciences* 14: 1715–1730.

Creutzfeldt B, Güntner A, Wziontek H, and Merz B (2010c) Reducing local hydrology from high precision gravity measurements: A lysimeter-based approach. *Geophysical Journal International* 183(1): 178–187. http://dx.doi.org/10.1111/j.1365-246X.2010.04742.x.

Creutzfeldt B, Kennedy J, and Ferré PA (2012b) Water-storage change measured with high-precision gravimetry at a groundwater recharge facility in Tucson, USA. In: *2012 AGU Fall Meeting*. San Francisco, CA: American Geophysical Union.

Creutzfeldt B, Troch PA, Ferré PA, Güntner A, Graeff T, and Merz B (2013) Storage-baseflow relationships at different catchment scales estimated by high-precision gravimetry. *Hydrological Processes,* Wiley online, http://dx.doi.org/10.1002/hyp.9689.

Creutzfeldt B, Troch PA, Güntner A, Ferré TPA, Graeff T, and Merz B (2014) Storage-discharge relationships at different catchment scales based on local high-precision gravimetry. *Hydrological Processes* 28(3): 1465–1475, Article first published online: 24. http://dx.doi.org/10.1002/hyp.9689.

Crossley D (1993) Core modes and Slichter modes – Fact and fancy. *Bulletin d'Informations des Marees Terrestres* 117: 8628–8638.

Crossley D (2004) Preface to the global geodynamics project. *Journal of Geodynamics* 38: 225–236.

Crossley D, de Linage C, Boy J-P, Hinderer J, and Famiglietti J (2012) A comparison of the gravity field over Central Europe from superconducting gravimeters, GRACE and global hydrological models, using EOF analysis. *Geophysical Journal International* 189: 877–897.

Crossley D and Hinderer J (2002) GGP ground truth for satellite gravity missions. *Bulletin d'Informations des Marees Terrestres* 136: 10735–10742.

Crossley D and Hinderer J (2005) Using SG arrays for hydrology in comparison with GRACE satellite data, with extension to seismic and volcanic hazards. *Korean Journal of Remote Sensing* 21(1): 31–49.

Crossley D and Hinderer J (2009) A review of the GGP network and scientific challenges. *Journal of Geodynamics* 48: 299–304.

Crossley D, Hinderer J, and Amalvict M (2001) A spectral comparison of absolute and superconducting gravimeter data. *Journal of the Geodetic Society of Japan* 47: 373–379.

Crossley D, Hinderer J, and Boy J-P (2004) Regional gravity variations in Europe from superconducting gravimeters. *Journal of Geodynamics* 38: 325–342.

Crossley D, Hinderer J, and Boy J-P (2005) Time variation of the European gravity field from superconducting gravimeters. *Geophysical Journal International* 161: 257–264.

Crossley D, Hinderer J, Casula G, et al. (1999) Network of superconducting gravimeters benefits a number of disciplines. *Eos, Transactions American Geophysical Union* 80: 121–126.

Crossley D, Hinderer J, Llubes M, and Florsch N (2003) The potential of ground gravity measurements to validate GRACE data. *Advances in Geosciences* 1: 1–7.

Crossley D, Hinderer J, and Riccardi U (2013) The measurement of surface gravity. *Reports on Progress in Physics* 76: 046101.

Crossley D, Jensen O, and Hinderer J (1995) Effective barometric admittance and gravity residuals. *Physics of the Earth and Planetary Interiors* 90: 221–241.

Crossley D, Jensen O, Xu H, and Hinderer J (1993) A slew rate detection criterion applied to SG data processing. *Bulletin d'Informations des Marees Terrestres* 117: 8675–8704.

Crossley D and Xu X (1998) Analysis of superconducting gravimeter data, from Table Mountain, Colorado. *Geophysical Journal International* 135: 835–844.

Crossley D, Xu H, and Van Dam T (1998) Comprehensive analysis of 2 years of data from Table Mountain, Colorado. In: Ducarme B and Paquet P (eds.) *Proceedings of the 13th International Symposium on Earth Tides, Brussels,* pp. 659–668.

Cummins P, Wahr J, Agnew D, and Tamura Y (1991) Constraining core undertones using stacked IDA gravity records. *Geophysical Journal International* 106: 189–198.

Dahlen FA and Sailor RV (1979) Rotational and elliptical splitting of the free oscillations of the Earth. *Geophysical Journal of the Royal Astronomical Society* 58: 609–623.

de Linage C, Hinderer J, and Boy J-P (2009a) Variability of the gravity-to-height ratio due to surface loads. *Pure and Applied Geophysics* 166: 1217–1245.

de Linage C, Hinderer J, and Rogister Y (2007) A search for the ratio between gravity variation and vertical displacement due to a surface load. *Geophysical Journal International* 171: 986–994.

de Linage C, Rivera L, Hinderer J, et al. (2009b) Separation of coseismic and postseismic gravity changes for the 2004 Sumatra-Andaman earthquake from 4.6 yr of GRACE observations and modelling of the coseismic change by normal-modes summation. *Geophysical Journal International* 176: 695–714.

de Meyer F and Ducarme B (1989) Non-tidal gravity changes observed with a superconducting gravimeter. In: *Proceedings of the 11th International Symposium on Earth Tides, Helsinki*, pp. 167–184. Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung.

Defraigne P, Dehant V, and Hinderer J (1994) Stacking gravity tide measurements and nutation observations in order to determine the complex eigenfrequency of the nearly diurnal free wobble. *Journal of Geophysical Research* 99(B5): 9203–9213.

Dehant V, Defraigne P, and Wahr J (1999) Tides for a convective Earth. *Journal of Geophysical Research* 104: 1035–1058.

Dickman SR (2005) Rotationally consistent Love numbers. *Geophysical Journal International* 161: 31–40.

Dierks O and Neumeyer J (2002) Comparison of Earth tides analysis programs. *Bulletin d'Informations des Marees Terrestres* 135: 10669–10688.

Döll P, Kaspar F, and Lehner B (2003) A global hydrological model for deriving water availability indicators: Model tuning and validation. *Journal of Hydrology* 270: 105–134.

Doodson AT (1921) The harmonic development of the tide generating potential. *Proceeding of the Royal Society of London* 100: 306–328.

Dragert H, Wang K, and James T (2001) A silent slip event on the deeper Cascadia subduction interface. *Science* 292: 1525–1528.

Dragert H, Wang K, and Rogers G (2004) Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth, Planets and Space* 56: 1143–1150.

Ducarme B (1983) Tidal gravity parameters at Brussels reconfirmed by a superconducting gravimeter. *Physics of the Earth and Planetary Interiors* 32: 1–3.

Ducarme B (2011) Determination of the main Lunar waves generated by the third degree tidal potential and validity of the corresponding body tides models. *Journal of Geodesy* 86(1): 65–75.

Ducarme B, Sun H-P, and Xu J-Q (2002) New investigation of tidal gravity results from the GGP network. *Bulletin d'Informations des Marees Terrestres* 136: 10761–10776.

Ducarme B, Venedikov A, Arnoso J, Chen X-D, Sun H-P, and Vieira R (2006) Global analysis of the GGP superconducting gravimeters network for the estimation of the pole tide gravimetric amplitude factor. *Journal of Geodynamics* 41: 334–344.

Ducarme B, Venedikov A, Arnoso J, and Vieira R (2004) Determination of the long period tidal waves in the GGP superconducting gravity data. *Journal of Geodynamics* 38: 307–324.

Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

Egbert GD and Erofeeva S (2002) Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology* 19: 183–204.

Elstner C and Schwahn W (1997) Precise mean parameters for daily and subdaily persistent air pressure waves at Potsdam for the period 1893–1992. In: Ducarme B and Paquet P (eds.) *Proceedings of the 13th International Symposium on Earth Tides, Brussels,* pp. 469–476.

Falk R, Harnisch M, Harnisch G, Nowak I, Richter B, and Wolf P (2001) Calibration of superconducting gravimeters SG103, C023, CD029, and CD030. *Journal of the Geodetic Society of Japan* 47(1): 22–27.

Farrell WE (1972) Deformation of the Earth by surface loads. *Reviews of Geophysics and Space Physics* 10(3): 751–797.

Flather RA (1976) A tidal model of the north-west European continental shelf. *Memoires Societe Royale des Sciences de Liege* 9: 141–164.

Florsch N and Hinderer J (2000) Bayesian estimation of the free core nutation parameters from the analysis of precise tidal gravity data. *Physics of the Earth and Planetary Interiors* 117: 21–35.

Florsch N, Hinderer J, and Legros H (1995a) Identification of quarter diurnal waves in superconducting gravimeter data. *Bulletin d'Informations des Marees Terrestres* 122: 9189–9198.

Florsch N, Legros H, and Hinderer J (1995b) The search for weak harmonic signals in a spectrum with application to gravity data. *Physics of the Earth and Planetary Interiors* 90: 197–210.

Francis O (1997) Calibration of the C021 superconducting gravimeter in Membach (Belgium) using 47 days of absolute gravity measurements. In: *Gravity, Geoid and Marine Geodesy, Tokyo, Japan*, IAG Symposium 117, pp. 212–219. Berlin: Springer.

Francis O and Hendrickx M (2001) Calibration of the LaCoste-Romberg 906 by comparison with the spring gravimeter C021 in Membach (Belgium). *Journal of the Geodetic Society of Japan* 47(1): 16–21.

Francis O, Lampitelli C, Klein G, Van Camp M, and Palinkas V (2011) Comparison between the transfer functions of three superconducting gravimeters. *Bulletin d'Informations des Marees Terrestres* 147: 11857–11868.

Francis O, Niebauer T, Sasagawa G, Klopping F, and Gschwind J (1998) Calibration of a superconducting gravimeter by comparison with an absolute gravimeter FG5 in Boulder. *Geophysical Research Letters* 25(7): 1075–1078.

Francis O, Van Camp M, van Dam T, Warnant R, and Hendrickx M (2004) Indication of the uplift of the Ardenne in long-term gravity variations in Membach (Belgium). *Geophysical Journal International* 158: 346–352.

Fratepietro F, Baker T, Williams SDP, and Van Camp M (2006) Ocean loading deformations caused by storm surges on the northwest European shelf. *Geophysical Research Letters* 33(6): L06317. http://dx.doi.org/10.1029/2005GL025475.

Freybourger M, Hinderer J, and Trampert J (1997) Comparative study of superconducting gravimeters and broadband seismometers STS-1/Z in subseismic frequency bands. *Physics of the Earth and Planetary Interiors* 101: 203–217.

Fukuda Y and Sato T (1997) Gravity effects of sea level variation at the superconducting gravimeter sites, estimated from ERS-1 and Topex-Poseidon altimeter data. In: Segawa J, Fujimoto H, and Okubo S (eds.) *Gravity, Geoid, and Marine Geodesy. IAG Symposia*, vol. 117, pp. 107–114. Berlin: Springer.

Furuya M, Okubo S, Sun W, Tanake Y, Oikawa J, and Watanabe H (2003) Spatiotemporal gravity changes at Miyakejima Volcano, Japan: Caldera collapse, explosive eruptions, and magma movement. *Journal of Geophysical Research* 108(B4): 2219.

Goodkind JM (1983) Q of the nearly diurnal free wobble. In: *Proceedings of the 9th International Symposium on Earth Tides, New York,* pp. 569–575.

Goodkind JM (1991) The superconducting gravimeter: Principles of operation, current performance and future prospects. In: *Proceedings of the Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Europeen de Geodynamique et de Seismologie, 3, Luxembourg,* pp. 81–90.

Goodkind JM (1996) Test of theoretical solid earth and ocean gravity tides. *Geophysical Journal International* 125: 106–114.

Goodkind JM (1999) The superconducting gravimeter. *Review of Scientific Instruments* 70(11): 4131–4152.

Goodkind JM, Czipott PV, Mills AP, et al. (1993) Test of the gravitational square law at 0.4 to 1.4 m mass separation. *Physical Review* 47(4): 1290–1297.

Goodkind JM and Young C (1991) Gravity and hydrology at Kilauea Volcano, the Geysers and Miami. In: Poitevin C (ed.) *Proceedings of the Workshop: Non Tidal Gravity Changes Intercomparison between absolute and superconducting gravimeters Cahiers Centre Europ. Géodyn. Séism. Lux. 3,* pp. 163–167.

Green J (1999) *Atmospheric Dynamics*. New York: Academic Press.

Guo J-Y, Dierks O, Neumeyer J, and Shum CK (2006) Weighting algorithms to stack superconducting gravimeter data for the potential detection of the Slichter modes. *Journal of Geodynamics* 41: 326–333.

Häfner R and Widmer-Schnidrig R (2012) Signature of 3-D density structure in spectra of the spheroidal free oscillation ${}_0S_2$. *Geophysical Journal International* 192: 285–294. http://dx.doi.org/10.1093/gji/ggs013.

Harnisch M and Harnisch G (1995) Processing of the data from two superconducting gravimeters, recorded in 1990–1991 at Richmond (Miami, Florida). Some problems and results. *Bulletin d'Informations des Marees Terrestres* 122: 9141–9147.

Harnisch M and Harnisch G (1997) Long time behaviour of superconducting gravimeters derived from observed time series. *Bulletin d'Informations des Marees Terrestres* 127: 9796–9805.

Harnisch M and Harnisch G (2002) Seasonal variations of hydrological influences on gravity measurements at Wettzell. *Bulletin d'Informations des Marees Terrestres* 137: 10.

Harnisch G and Harnisch M (2006a) Hydrological influences in long gravimetric data series. *Journal of Geodynamics* 41: 276–287.

Harnisch M and Harnisch G (2006b) Study of long-term gravity variations, based on data of the GGP co-operation. *Journal of Geodynamics* 41(1–3): 318–325.

Harnisch M, Harnisch G, and Falk R (2002) Improved scale factors of the BKG superconducting gravimeters, derived from comparisons with absolute gravity measurements. *Bulletin d'Informations des Marees Terrestres* 135: 10627–10642.

Harnisch M, Harnisch G, Nowak I, Richter B, and Wolf P (2000) The dual sphere superconducting gravimeter C029 at Frankfurt a.M. and Wettzell. First results and calibration. In: *Proceedings of the Workshop: High-Precision Gravity Measurements with Application to Geodynamics and Second GGP Workshop, Cahiers du Centre Européen de Géodynamique et de Séismologie, 17, Luxembourg,* pp. 39–56.

Harnisch M, Harnisch G, Richter B, and Schwahn W (1998) Estimation of polar motion effects from time series recorded by superconducting gravimeters. In: *Proceedings of the 13th International Symposium on Earth Tides, Brussels 1997*, pp. 511–518. Brussels: Observatoire Royal de Belgique.

Hartmann T and Wenzel H-G (1995a) The HW95 tidal potential catalogue. *Geophysical Research Letters* 22(24): 3553–3556.

Hartmann T and Wenzel H-G (1995b) Catalog HW95 of the tide generating potential. *Bulletin d'Informations des Marees Terrestres* 123: 9278–9301.

Hasan S, Troch PA, Bogaart PW, and Kroner C (2008) Evaluating catchment-scale hydrological modeling by means of terrestrial gravity observations. *Water Resources Research* 44: W08416. http://dx.doi.org/10.1029/2007WR006321.

Hasan S, Troch PA, Boll J, and Kroner C (2006) Modeling the hydrological effect in local gravity at Moxa, Germany. *Journal of Hydrometeorology* 7(3): 346–354. http://dx.doi.org/10.1175/JHM488.1.

Herring TA, Gwinn CR, and Shapiro II (1986) Geodesy by radio Interferometry: Studies of the forced nutations of the Earth: I. Data analysis. *Journal of Geophysical Research* 91(B5): 4745–4754.

Hinderer J, Amalvict M, Crossley D, Rivera L, Leveque J-J, and Luck B (2002a) Tides, earthquakes and ground noise as seen by an absolute gravimeter and its superspring; A comparison with a broadband seismometer and a superconducting gravimeter. *Metrologia* 39: 495–501.

Hinderer J, Andersen O, Lemoine F, Crossley D, and Boy J-P (2006) Seasonal changes in the European gravity field from GRACE: A comparison with superconducting gravimeters and hydrology model predictions. *Journal of Geodynamics* 41: 59–68.

Hinderer J, Boy JP, Gegout P, Defraigne P, Roosbeek F, and Dehant V (2000) Are the free core nutation parameters variable in time? *Physics of the Earth and Planetary Interiors* 117: 37–49.

Hinderer J, Boy JP, and Legros H (1998) A 3000 day registration of the superconducting gravimeter GWR T005 in Strasbourg, France. In: Ducarme B and Paquet P (eds.) *Proceedings of the 13th International Symposium on Earth Tides,* pp. 617–624.

Hinderer J and Crossley D (2000) Time variations in gravity and inferences on the Earth's structure and dynamics. *Surveys in Geophysics* 21: 1–45.

Hinderer J and Crossley D (2004) Scientific achievements from the first phase (1997–2003) of the Global Geodynamics Project using a worldwide network of superconducting gravimeters. *Journal of Geodynamics* 38: 237–262.

Hinderer J, Crossley D, and Jensen O (1995) A search for the Slichter triplet in superconducting gravimeter data. *Physics of the Earth and Planetary Interiors* 90: 183–195.

Hinderer J, Crossley D, and Xu H (1994) A two year comparison between the French and Canadian superconducting gravimeter data. *Geophysical Journal International* 116: 252–266.

Hinderer J, Florsch N, Mäkinen J, Legros H, and Faller JE (1991) On the calibration of a superconducting gravimeter using absolute gravity measurements. *Geophysical Journal International* 106: 491–497.

Hinderer J and Legros H (1989) Gravity perturbations of annual period. In: *Proceedings of the 11th International Symposium on Earth Tides*, Helsinki, pp. 425–429. Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung.

Hinderer J, Pfeffer J, Boucher M, et al. (2012) Water storage changes from ground and space geodesy: First results from the GHYRAF (gravity and hydrology in Africa) experiment. *Pure and Applied Geophysics* 169(8): 1391–1410. http://dx.doi.org/10.1007/s00024-011-0417-9.

Hinderer J, Rosat S, Crossley D, Amalvict M, Boy J-P, and Gegout P (2002b) Influence of different processing methods on the retrieval of gravity signals from GGP data. *Bulletin d'Informations des Marees Terrestres* 123: 9278–9301.

Hirt C, Kuhn M, Featherstone WE, and Gottl F (2012) Topographic/isostatic evaluation of new-generation GOCE gravity field models. *Journal of Geophysics Research* 117: B05407.

Hsu H-T, Tao GQ, Song XL, Baker TF, Edge RJ, and Jeffries G (1989) Gravity tidal datum at Wuchang of China. In: *Proceedings of the 11th International Symposium on Earth Tides, Helsinki*, pp. 187–195. Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung.

Ijpelaar R, Troch P, Warderdam P, Stricker H, and Ducarme B (2002) Detecting hydrological signals in time series of in-situ gravity measurements. *Bulletin d'Informations des Marees Terrestres* 135: 10837–10838.

Imanishi Y (2005) On the possible cause of long period instrumental noise (parasitic mode) of a superconducting gravimeter. *Journal of Geodesy* 78(11–12): 683–690.

Imanishi Y, Higashi T, and Fukuda Y (2002) Calibration of the superconducting gravimeter T011 by parallel observation with the absolute gravimeter FG5 210 – A Bayesian approach. *Geophysical Journal International* 151: 867–878.

Imanishi Y, Kokubo K, and Tatehata H (2006) Effect of underground water on gravity observation at Matsushiro, Japan. *Journal of Geodynamics* 41: 221–226.

Imanishi Y, Sato T, Higashi T, Sun W, and Okubo S (2004) A network of superconducting gravimeters detects submicrogal coseismic gravity changes. *Science* 306: 476–478.

Imanishi Y, Tamura Y, Ikeda H, and Okubo S (2007) Permanent gravity changes recorded on superconducting gravimeters from earthquakes in central Japan – The Noto Hantou and Niigataken Chuetsu-oki events in 2007. *Journal of Geodynamics* 48: 260–268.

Iwano S and Fukuda Y (2004) Superconducting gravimeter observations without a tilt compensation system. *Physics of the Earth and Planetary Interiors* 147(4): 343–351.

Iwano S, Kimura I, and Fukuda Y (2003) Calibration of the superconducting gravimeter TT70 #016 at Syowa station by parallel observation with the absolute gravimeter FG5 #203. *Polar Geoscience* 16: 22–28.

Jacob T, Bayer R, Chery J, et al. (2008) Absolute gravity monitoring of water storage variation in a karst aquifer on the larzac plateau (Southern France). *Journal of Hydrology* 359(12): 105–117.

Jacob T, Chery J, Bayer R, et al. (2009) Time-lapse surface to depth gravity measurements on a karst system reveal the dominant role of the epikarst as a water storage entity. *Geophysical Journal International* 177: 347–360.

James TS and Ivins ER (1998) Predictions of Antarctic crustal motions driven by present-day ice sheet evolution and by isostatic memory of the Last Glacial Maximum. *Journal of Geophysical Research* 103: 4993–5017.

Jentzsch G, Kroner C, Flach D, and Gommlich G (1995) Long and aperiodic effects in the recording of the superconducting gravimeter in the Asse salt mine in Northern Germany. In: *Proceedings of the 2nd Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Européen de Géodynamique et de Séismologie, 11, Luxembourg,* pp. 187–189.

Kennedy J, Ferré TPA, Güntner A, Abe M, and Creutzfeldt B (2014) Direct measurement of subsurface mass change using the variable baseline gravity gradient method. *Geophysical Research Letters* 41: 2827–2834. http://dx.doi.org/10.1002/2014GL059673.

Klopping FJ, Peter G, Berstis KA, Carter WE, Goodkind JM, and Richter BD (1995) Analysis of two 525 day long data sets obtained with two side-by-side, simultaneously recording superconducting gravimeters at Richmond, Florida, USA. In: *Proceedings of the 2nd Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Europeen de Geodynamique et de Seismologie, 11, Luxembourg,* pp. 57–69.

Klügel T and Wziontek H (2009) Correcting gravimeters and tiltmeters for atmospheric mass attraction using operational weather models. *Journal of Geodynamics* 48: 204–210.

Kobayashi N and Nishida K (1998) Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature* 395: 357–360.

Kroner C (2001) Hydrological effects on gravity at the Geodynamic Observatory Moxa. *Journal of Geodetic Society of Japan* 47(1): 353–358.

Kroner C, Dierks O, Neumeyer J, and Wilmes H (2005) Analysis of observations with dual sensor superconducting gravimeters. *Physics of Earth and Planetary Interiors* 153: 210–219.

Kroner C, Dobslaw H, Thomas M, Janssen F, and Weise A (2009) Non-tidal oceanic mass shifts and their impact on terrestrial gravity observations. *Geophysical Research Abstracts* 11, EGU2009-4079.

Kroner C and Jahr T (2006) Hydrological experiments around the superconducting gravimeter at Moxa Observatory. *Journal of Geodynamics* 41: 268–275.

Kroner C, Jahr T, and Jentzsch G (2004) Results from 44 months of observations with a superconducting gravimeter at Moxa/Germany. *Journal of Geodynamics* 38(3–5): 263–280.

Kroner C, Jahr T, Naujoks M, and Weise A (2007) Hydrological signals in gravity – Foe or friend? *Dynamic Planet. IAG Symposia*, vol. 130, pp. 504–510. Berlin/Heidelberg: Springer.

Kroner C and Jentzsch G (1999) Comparison of different pressure reductions for gravity data and resulting consequences. *Physics of the Earth and Planetary Interiors* 115: 205–218.

Kudryavtsev SM (2004) Improved harmonic development of the Earth tide-generating potential. *Journal of Geodesy* 77: 829–838.

Lambert A (1974) Earth tide analysis and prediction by the response method. *Journal of Geophysical Research* 79(32): 4952–4960.

Lambert A, Courtier N, and James TS (2006) Long-term monitoring by absolute gravimetry: Tides to postglacial rebound. *Journal of Geodynamics* 41: 307–317.

Lambert A, Pagiatakis SD, Billyard AP, and Dragert H (2003) Improved ocean tidal loading corrections for gravity and displacement: Canada and northern United States. *Journal of Geophysical Research* 103(B12): 30231–30244.

Larson JV and Harrison JC (1986) An improved analysis of the electrostatic feedback of LaCoste and Romberg gravity meters. In: *Proceedings of the 10th International Symposium on Earth Tides, Madrid 1985*, pp. 1–8. Madrid: Consejo Superior de Investigaciones Cientificas.

Lefèvre F, Lyard F, Le Provost C, and Schrama EJO (2002) FES99: A global tide finite element solution assimilating tide gauge and altimetric information. *Journal of Atmospheric and Oceanic Technology* 19: 1345–1356.

Lognonne P, Clevede E, and Kanamori H (1998) Computation of seismograms and atmospheric oscillations by normal-mode summation for a spherical earth model with realistic atmosphere. *Geophysical Journal International* 135: 388–406.

Longuevergne L, Boy J-P, Florsch N, et al. (2009) Local and global hydrological contributions to gravity variations observed in Strasbourg. *Journal of Geodynamics* 48(3–5): 189–194, 11.

Loyer S, Hinderer J, and Boy JP (1999) Determination of the gravimetric factor at the Chandler period from Earth's orientation data and superconducting gravimetry observations. *Geophysical Journal International* 136: 1–7.

Lynch DR and Gray WG (1979) A wave equation model for finite element tidal computations. *Computers and Fluids* 7: 207–228.

Mathews PM (2001) Love numbers and gravimetric factor for diurnal tides. *Journal of the Geodetic Society of Japan* 47: 231–236.

Mathews PM, Herring TA, and Buffett B (2002) Modeling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior. *Journal of Geophysical Research* 107(B4). http://dx.doi.org/10.1029/2001JB000390.

Matsumoto K, Sato T, Takanezawa T, and Ooe M (2001) GOTIC2: A program for the computation of ocean tidal loading effect. *Journal of the Geodetic Society of Japan* 47: 243–248.

Matsumoto K, Takanezawa T, and Ooe M (2000) Ocean tide models developed by assimilating TOPEX/Poseidon altimeter data into hydrodynamical model: A global model and a regional model around Japan. *Journal of Oceanography* 56: 567–581.

Matsuo K and Heki K (2011) Coseismic gravity changes of the 2011 Tohoku-oki earthquake from satellite gravimetry. *Geophysical Research Letters* 38, L00G12.

Mazzotti S, Lambert A, Courtier N, Nykolaishen L, and Dragert H (2007) Crustal uplift and sea level rise in northern Cascadia from GPS, absolute gravity, and tide gauge data Geophys. *Research Letters* 34: L15306. http://dx.doi.org/10.1029/2007GL030283.

Melchior P and Ducarme B (1986) Detection of inertial gravity oscillations in the Earth's core with a superconducting gravimeter at Brussels. *Physics of the Earth and Planetary Interiors* 42: 129–134.

Melchior P and Georis B (1968) Earth tides, precession-nutations and the secular retardation of Earth's rotation. *Physics of the Earth and Planetary Interiors* 1(4): 267–287.

Mémin A, Hinderer J, and Rogister Y (2012) Separation of the geodetic consequences of past and present ice-mass change: Influence of topography with application to

Svalbard (Norway). *Pure and Applied Geophysics* 169(8): 1357–1372. http://dx.doi.org/10.1007/s00024-011-0399-7.

Mémin A, Rogister Y, Hinderer J, Omang OC, and Luck B (2011) Secular gravity variation at Svalbard (Norway) from ground observations and GRACE satellite data. *Geophysical Journal International* 184: 1119–1130.

Merlet S, Kopaev A, Diament M, Geneves G, Landragin A, and Pereira Dos Santos F (2008) Micro-gravity investigations for the LNE watt balance project. *Metrologia* 45: 265–274.

Merriam J (1992a) An ephemeris for gravity tide predictions at the 1 ngal level. *Geophysical Journal International* 108: 415–422.

Merriam J (1992b) Atmospheric pressure and gravity. *Geophysical Journal International* 109: 488–500.

Merriam J (1993) A comparison of recent tide catalogues. *Bulletin d'Informations des Marees Terrestres* 115: 8515–8535.

Merriam J (1994) The nearly diurnal free wobble resonance in gravity measured at Cantley, Quebec. *Geophysical Journal International* 119: 369–380.

Merriam J (1995a) Calibration, phase stability, and a search for non-linear effects in data from the superconducting gravimeter at Cantley, Quebec. In: Hsu H-T (ed.) *Proceedings of the 12th International Symposium on Earth Tides, Beijing, China*, pp. 311–316. New York: Science Press.

Merriam J (1995b) Non-linear tides observed with the superconducting gravimeter at Cantley, Quebec. *Geophysical Journal International* 123: 529–540.

Merriam J (2000) The response method applied to the analysis of superconducting gravimeter data. *Physics of the Earth and Planetary Interiors* 121(3–4): 289–299.

Merriam J, Pagiatakis S, and Liard J (2001) Reference level stability of the Canadian superconducting gravimeter installation. *Journal of the Geodetic Society of Japan* 47(1): 417–423.

Meurers B (1999) Air pressure signatures in the SG data of Vienna. *Bulletin d'Informations des Marees Terrestres* 131: 10195–10200.

Meurers B (2000) Gravitational effects of atmospheric processes in SG gravity data. In: Ducarme B and Barthelemy J (eds.) *Proceedings the Workshop: 'High Precision Gravity Measurements with Application to Geodynamics and Second GGP Workshop,' Luxembourg, 1999, ECGS Cahiers, 17,* pp. 57–65.

Meurers B (2001a) Superconducting gravimetry in Geophysical research today. *Journal of the Geodetic Society of Japan* 47(1): 300–307.

Meurers B (2001b) Tidal and non-tidal gravity variations in Vienna – A Five Years' SG record. *Journal of the Geodetic Society of Japan* 47(1): 392–397.

Meurers B (2002) Aspects of gravimeter calibration by time domain comparison of gravity records. *Bulletin d'Informations des Marees Terrestres* 135: 10643–10650.

Meurers B (2004) Investigation of temporal gravity variations in SG-records. *Journal of Geodynamics* 38: 423–435.

Meurers B (2012) Superconducting gravimeter calibration by colocated gravity observations: Results from GWR C025. *International Journal of Geophysics* 2012: 12. http://dx.doi.org/10.1155/2012/954271, Article ID 954271.

Meurers B, Van Camp M, and Petermans T (2007) Correcting superconducting gravity time-series using rainfall modelling at the Vienna and Membach stations and application to Earth tide analysis. *Journal of Geodesy* 81: 703–712. http://dx.doi.org/10.1007/s00190-007-0137-1.

Mikhailov V, Tikhotsky S, Diament M, Panet I, and Ballu V (2004) Can tectonic processes be recovered from new gravity satellite data? *Earth and Planetary Science Letters* 228(3–4): 281–297.

Milly C and Shmakin A (2002) Global modeling of land water and energy balances. Part I: The land dynamics (LaD) model. *Journal of Hydrometeorology* 3: 283–299.

Mouyen M, Masson F, Hwang C, et al. (2009) Expected temporal absolute gravity change across the Taiwanese Orogen, a modeling approach. *Journal of Geodynamics* 48(3–5): 284–291.

Mouyen M, Masson F, Hwang C, et al. (2012) Erosion effects assessed by repeated gravity measurements in southern Taiwan. *Geophysical Journal International*, 192: 113–136. http://dx.doi.org/10.1093/gji/ggs019.

Mukai A, Higashi T, Takemoto S, Nakagawa I, and Naito I (1995) Accurate estimation of atmospheric effects on gravity observations made with a superconducting gravimeter at Kyoto. *Physics of the Earth and Planetary Interiors* 91: 149–159.

Mukai A, Takemoto S, Higashi T, and Fukuda Y (2001) Oceanic tidal loadings estimated from gravity observations in Kyoto and Bandung. *Journal of the Geodetic Society of Japan* 47(1): 261–266.

Müller T and Zürn W (1983) Observation of gravity changes during the passage of cold fronts. *Journal of Geophysics* 53: 155–162.

Naujoks M, Kroner C, Weise W, Jahr T, Krause P, and Eisner S (2010) Evaluating local hydrological modelling by temporal gravity observations and a gravimetric three-dimensional model. *Geophysical Journal International* 182(1): 233–249. http://dx.doi.org/10.1111/j.1365-246X.2010.04615.x.

Nawa K, Suda N, Fukao Y, Sato T, Aoyama Y, and Shibuya K (1998) Incessant excitation of the Earth's free oscillations. *Earth, Planets and Space* 50: 3–8.

Nawa K, Suda N, Fukao Y, et al. (2000) Incessant excitation of the Earth's free oscillations: Global comparison of superconducting gravimeter records. *Physics of the Earth and Planetary Interiors* 120: 289–297.

Nawa K, Suda N, Yamada I, Miyajima R, and Okubo S (2009) Coseismic change and precipitation effect in temporal gravity variation at Inuyama, Japan: A case of the 2004 off the Kii peninsula earthquakes observed with a superconducting gravimeter. *Journal of Geodynamics* 48(1): 1–5.

Neuberg J, Hinderer J, and Zürn W (1987) Stacking gravity tide observations in Central Europe for the retrieval of the complex eigenfrequency of the nearly diurnal free wobble. *Geophysical Journal of the Royal Astronomical Society* 91: 853–868.

Neumeyer J (1995) Frequency-dependent atmospheric pressure correction on gravity variations by means of cross-spectral analysis. *Bulletin d'Informations des Marees Terrestres* 122: 9212–9220.

Neumeyer J (2010) Superconducting gravimetry. In: Xu G (ed.) *Sciences of Geodesy – I*, pp. 339–413. Berlin Heidelberg: Springer.

Neumeyer J, Barthelmes F, Dierks O, et al. (2006) Combination of temporal gravity variations resulting from superconducting gravimeter (SG) recordings, GRACE satellite observations and global hydrology models. *Journal of Geodesy* 79: 573–585. http://dx.doi.org/10.1007/s00190-005-0014-8.

Neumeyer J, Barthelmes B, Kroner C, et al. (2008) Analysis of gravity field variations derived from superconducting gravimeter recordings, the GRACE satellite and hydrological models at selected European sites. *Earth, Planets and Space* 60: 505–518.

Neumeyer J, Hagedoorn J, Leitloff J, and Schmidt R (2004a) Gravity reduction with three-dimensional atmospheric pressure data for precise ground gravity measurements. *Journal of Geodynamics* 38: 437–450.

Neumeyer J, Schwintzer P, Barthelmes F, et al. (2004b) Comparison of superconducting gravimeter and CHAMP satellite derived temporal gravity variations. In: Reigber Ch, Lühr H, Schwintzer P, and Wickert J (eds.) *Earth Observations with CHAMP Results from Three Years in Orbit*, pp. 31–36.

Neumeyer J and Stobie B (2000) The new superconducting gravimeter site at the South African geodynamic observatory Sutherland (SAGOS). In: *Proceedings of the Workshop: High-Precision Gravity Measurements with Application to Geodynamics and Second GGP Workshop, Cahiers du Centre Européen de Géodynamique et de Séismologie, 17, Luxembourg,* pp. 85–96.

Niebauer T, Sasagawa G, Faller J, Hilt R, and Klopping F (1995) A new generation of absolute gravimeters. *Metrologia* 32: 159–180.

Nishida K and Kobayashi N (1999) Statistical features of Earth's continuous free oscillations. *Journal of Geophysical Research* 104: 28741–28750.

Nishida K, Kobayashi N, and Fukao Y (2000) Resonant oscillations between the solid Earth and the atmosphere. *Science* 287: 2244–2246.

Nishida K, Kobayashi N, and Fukao Y (2002) Origin of Earth's ground noise from 2 to 20 mHz. *Geophysical Research Letters* 29: 52-1–52-4.

Oka D, Fujimitsu Y, Nishijima J, Fukuda Y, and Taniguchi M (2012) Evaluation of geothermal reservoir mass change from the gravity change at the Takigami geothermal area, Oita prefecture, Japan. In: *Proceedings of the Thirty-Seventh Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, California, January 30–February 1, 2012,* SGP-TR-194.

Okubo S, Yoshida S, Sato T, Tamura Y, and Imanishi Y (1997) Verifying the precision of an absolute gravimeter FG5: Comparison with superconducting gravimeters and detection of oceanic tidal loading. *Geophysical Research Letters* 24: 489–492.

Olson JJ and Warburton RJ (1979) Continuous gravity observations at The Geysers: A preliminary report. *Geothermal Resources Council Transactions* 3: 519–522.

Omang OC and Kierulf HP (2011) Past and present-day ice mass variation on Svalbard revealed by superconducting gravimeter and GPS measurements. *Geophysical Research Letters* 38: L22304.

Pagiatakis SD (2000) Superconducting gravimeter data treatment and analysis. *Cahiers du Centre Européen de Géodynamique et Séismologie* 17: 103–113.

Panet I, Mikhailov V, Diament M, et al. (2007) Co-seismic and post-seismic signatures of the Sumatra December 2004 and March 2005 earthquakes in GRACE satellite gravity. *Geophysical Journal International* 171(1): 177–190.

Peter G, Klopping F, and Berstis KA (1995) Observing and modeling gravity changes caused by soil moisture and groundwater table variations with superconducting gravimeters in Richmond, Florida. In: *Cahiers du Centre Europeen G'eodynamique et de S'eismologie Luxembourg,* pp. 147–159.

Peterson J (1993) Observations and modelling of seismic background noise. *Open-File Report 93-332*. Albuquerque, New Mexico: US Department of Interior, Geological Survey.

Pfeffer J, Boucher M, Hinderer J, et al. (2011) Hydrological contribution to time – Variable gravity: Influence of the West African Monsoon in Southwest Niger. *Geophysical Journal International* 184(2): 661–672.

Pingree RD and Griffiths KD (1980) Currents driven by a steady uniform wind stress on the shelf seas around the British Isles. *Oceanologica Acta* 3: 227–235.
Pingree RD and Griffiths KD (1981) S_2 tidal simulations on the North-West European Shelf. *Journal of the Marine Biological Association of the UK* 61: 609–616.
Prothero WA (1967) *A Cryogenic Gravimeter*. PhD Thesis, University of California at San Diego, La Jolla.
Prothero WA and Goodkind JM (1968) A superconducting gravimeter. *Review of Scientific Instruments* 39: 1257–1262.
Prothero WA and Goodkind JM (1972) Earth tide measurements with the superconducting gravimeter. *Journal of Geophysical Research* 77: 926–932.
Rabbell W and Zschau J (1985) Static deformations and gravity changes at the earth's surface due to atmospheric loading. *Journal of Geophysics* 56: 81–99.
Ray RD (1999) *A Global Ocean Tide Model from TOPEX/Poseidon Altimeter: GOT99.2, NASA Tech. Memo., TM-209478*, 58 pp.
Rhie J and Romanowicz B (2004) Excitation of Earth's continuous free oscillations by atmosphere–ocean–seafloor coupling. *Nature* 431. http://dx.doi.org/10.1038/nature02942.
Riccardi U, Berrino G, Corrado G, and Hinderer J (2008) Strategies in the processing and analyses of continuous gravity record in active volcanic areas: The case of Mt. Vesuvius. *Annals of Geophysics* 51(1): 67–85.
Riccardi U, Hinderer J, and Boy J-P (2007) On the efficiency of barometric arrays to improve the reduction of atmospheric effects on gravity data. *Physics of the Earth and Planetary Interiors* 161(3–4): 224–242.
Riccardi U, Hinderer J, Boy J-P, and Rogister Y (2009) Tilt effects on GWR superconducting gravimeters. *Journal of Geodynamics* 48: 316–324.
Riccardi U, Rosat S, and Hinderer J (2011) A 300-day parallel gravity record with the gPhone-054 spring gravimeter and the GWR-C026 superconducting gravimeter in Strasbourg (France): First comparative study. *Metrologia* 48(1): 28–39.
Riccardi U, Rosat S, and Hinderer J (2012) On the accuracy of the calibration of superconducting gravimeters using absolute and spring sensors: A critical comparison. *Pure and Applied Geophysics* 169: 1343–1356.
Richter B (1983) The long-period tides in the Earth tide spectrum. In: *Proceedings XVIII General Assembly IAG, Hamburg 1*, pp. 204–216. Columbus, OH: Ohio State University Press, 1984.
Richter B (1985) Three years of registration with the superconducting gravimeter. *Bulletin d'Informations des Marees Terrestres* 94: 6344–6352.
Richter B (1987) *Das supraleitende Gravimeter*. Deutsche Geodät. Komm., Reihe C, 329, Frankfurt am Main, 124 pp.
Richter B (1990) The long period elastic behavior of the Earth. In: McCarthy D and Carter W (eds.) *IUGG Geophysical Monograph No. 59,* 9, pp. 21–25.
Richter B (1991) Calibration of superconducting gravimeters. In: *Proceedings of the Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Europeen de Geodynamique et de Seismologie, 3, Luxembourg,* pp. 99–107.
Richter B and Warburton RJ (1998) A new generation of superconducting gravimeters. In: *Proceedings of the 13th International Symposium Earth Tides, Brussels 1997*, pp. 545–555. Brussels: Observatoire Royal de Belgique.
Richter B and Wenzel H-G (1991) Precise instrumental phase lag determination by the step response method. *Bulletin d'Informations des Marees Terrestres* 111: 8032–8052.
Richter B, Wenzel H-G, Zürn W, and Klopping F (1995a) From Chandler wobble to free oscillations: Comparison of cryogenic gravimeters and other instruments over a wide period range. *Physics of the Earth and Planetary Interiors* 91: 131–148.
Richter B, Wilmes H, and Nowak I (1995b) The Frankfurt calibration system for relative gravimeters. *Metrologia* 32: 217–223.
Richter B, Zerbini S, Matonti F, and Simon D (2004) Long-term crustal deformation monitored by gravity and space techniques at Medicina, Italy and Wettzell, Germany. *Journal of Geodynamics* 38: 281–292.
Rieutord M (2002) Slichter modes of the Earth revisited. *Physics of the Earth and Planetary Interiors* 131: 269–278.
Rochester MG and Crossley DJ (2009) Earth's long-period wobble modes: A Lagrangian description. *Geophysical Journal International* 176: 40–62.
Rodell M and Famiglietti J (1999) Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. *Water Resources Research* 35(9): 2705–2723.
Rodell M, Houser PR, Jambor U, et al. (2004) The global land data assimilation system. *Bulletin of the American Meteorological Society* 85(3): 381–394.
Rogers G and Dragert H (2003) Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* 300(5627): 1942. http://dx.doi.org/10.1126/science.1084783.
Romagnoli C, Zerbini S, Lago L, et al. (2003) Influence of soil consolidation and thermal expansion effects on height and gravity variations. *Journal of Geodynamics* 35: 521–539.
Roosbeek F (1996) RATGP95: A harmonic development of the tide-generating potential using an analytic method. *Geophysical Journal International* 126: 197–204.
Rosat S, Boy J-P, Ferhat G, et al. (2009a) Analysis of a ten-year (1997–2007) record of time-varying gravity in Strasbourg using absolute and superconducting gravimeters: New results on the calibration and comparison with GPS height changes and hydrology. *Journal of Geodynamics* 48(3–5): 360–365.
Rosat S, Florsch N, Hinderer J, and Llubes M (2009b) Estimation of the free core nutation parameters from SG data: Sensitivity study and comparative analysis using linearized least-squares and Bayesian methods. *Journal of Geodynamics* 48: 331–339.
Rosat S and Hinderer J (2011) Noise levels of superconducting gravimeters: Updated comparison and time stability. *Bulletin of the Seismological Society of America* 101(3): 1233–1241. http://dx.doi.org/10.1785/0120100217.
Rosat S, Hinderer J, Crossley D, and Boy J-P (2004) Performance of superconducting gravimeters from long-period seismology to tides. *Journal of Geodynamics* 38: 461–476.
Rosat S, Hinderer J, Crossley D, and Rivera L (2003a) The search for the Slichter Mode: Comparison of noise levels of superconducting gravimeters and investigation of a stacking method. *Physics of the Earth and Planetary Interiors* 140: 183–202.
Rosat S, Hinderer J, and Rivera L (2003b) First observation of ${}_2S_1$ and study of the splitting of the football mode ${}_0S_2$ after the June 2001 Peru event of magnitude 8.4. *Geophysical Research Letters* 30(21): 2111. http://dx.doi.org/10.1029/2003GL018304.
Rosat S and Rogister Y (2012) Excitation of the Slichter mode by collision with a meteoroid or pressure variations at the surface and core boundaries. *Physics of the Earth and Planetary Interiors* 190–191: 25–33.
Rosat S, Rogister Y, Crossley D, and Hinderer J (2006) A search for the Slichter triplet with superconducting gravimeters: Impact on the density jump at the inner core boundary. *Journal of Geodynamics* 41: 296–306.
Rosat S, Sato T, Imanishi Y, et al. (2005) High-resolution analysis of the gravest seismic normal modes after the 2004 $M_w = 9$ Sumatra earthquake using superconducting gravimeter data. *Geophysical Research Letters* 32: L13304. http://dx.doi.org/10.1029/2005GL023128.
Rosat S, Watada S, and Sato T (2007) Geographical variations of the ${}_0S_0$ normal mode amplitude: Predictions and observations after the Sumatra-Andaman earthquake. *Earth, Planets and Space* 59: 307–311.
Roult G and Crawford W (2000) Analysis of 'background' free oscillations and how to improve resolution by subtracting the atmospheric pressure signal. *Physics of the Earth and Planetary Interiors* 121: 325–338.
Rymer H and Locke C (1995) Microgravity and ground deformation precursors to eruption: A review. In: *Proceedings of the Workshop: New Challenges for Geodesy in Volcano Monitoring. Cah. Cent. Europ. Géodyn. et Séism., Luxembourg,* pp. 21–39.
Sato T, Boy J-P, Tamura Y, et al. (2006a) Gravity tide and seasonal gravity variation at Ny-Ålesund, Svalbard in Arctic. *Journal of Geodynamics* 41: 234–241.
Sato T, Fukuda Y, Aoyama Y, et al. (2001) On the observed annual gravity variation and the effect of sea surface height variations. *Physics of the Earth and Planetary Interiors* 123: 45–63.
Sato T, Miura S, Sun W, et al. (2012) Gravity and uplift rates observed in southeast Alaska and their comparison with GIA model predictions. *Journal of Geophysical Research* 117: B01401. http://dx.doi.org/10.1029/2011JB008485.
Sato T, Nawa K, Shibuya K, et al. (1997a) Polar motion effect on gravity observed with a superconducting gravimeter at Syowa station, Antarctica. In: Segawa J, et al. (eds.) *Gravity, Geoid, and Marine Geodesy. IAG Symposia*, vol. 117, pp. 99–106. Berlin: Springer.
Sato T, Okuno J, Hinderer J, et al. (2006b) A geophysical interpretation of the secular displacement and gravity rates observed at Ny-Alesund, Svalbard in the Arctic – Effects of the post-glacial rebound and present-day ice melting. *Geophysical Journal International* 165: 729–743.
Sato T, Ooe M, Nawa K, Shibuya K, Tamura Y, and Kaminuma K (1997b) Long-period tides observed with a superconducting gravimeter at Syowa station, Antarctica, and their implication to global ocean tide modeling. *Physics of the Earth and Planetary Interiors* 103: 39–53.
Sato T, Shibuya K, Ooe M, et al. (1991) Long term stability of the superconducting gravimeter installed at Syowa station, Antarctica. In: *Proceedings of the 2nd Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Européen de Géodynamique et de Séismologie, Luxembourg, 11,* pp. 71–75.
Sato T, Tamura Y, Matsumoto K, Imanishi Y, and McQueen H (2002) Parameters of the fluid core resonance estimated from superconducting gravimeter data. *Bulletin d'Informations des Marees Terrestres* 136: 10751–10760.

Sato T, Tamura Y, Matsumoto K, Imanishi Y, and McQueen H (2004) Parameters of the fluid core resonance inferred from superconducting gravimeter data. *Journal of Geodynamics* 38: 375–389.

Scherneck H-G (1991) A parameterized Earth tide observation model and ocean tide loading effects for precise geodetic measurements. *Geophysical Journal International* 106: 677–695.

Schwahn W, Baker T, and Falk R (2000) Long-term increase of gravity at the medicina station (Northern Italy) confirmed by absolute and superconducting gravimetric time series. In: *Proceedings of the Workshop: High-Precision Gravity Measurements with Application to Geodynamics and Second GGP Workshop, Cahiers du Centre Européen de Géodynamique et de Séismologie, 17, Luxembourg,* pp. 145–168.

Schwiderski EW (1980) On charting global ocean tides. *Reviews of Geophysics* 18: 243–268.

Seama N, Fukuda Y, and Segawa J (1993) Superconducting gravimeter observations at Kakioka, Japan. *Journal of Geomagnetism and Geoelectricity* 45: 1383–1394.

Shiomi S (2008a) Proposal for a geophysical search for dilatonic waves. *Physical Review D* 78, 042001–042006.

Shiomi S (2008b) Testing gravitational physics with superconducting gravimeters. *Progress of Theoretical Physics* supplement 172: 61–70.

Simon D (2002) Modeling of the field of gravity variations induced by the seasonal air mass warming during 1998–2000. *Bulletin d'Informations des Marees Terrestres* 136: 10821–10836.

Sinha B and Pingree RD (1997) The principal lunar semidiurnal tide and its harmonics: Baseline solutions for M_2 and M_4 constituents on the North-West European continental shelf. *Continental Shelf Research* 17: 1321–1365.

Smylie DE (1992) The inner core translational triplet and the density near Earth's center. *Science* 255: 1678–1682.

Smylie D, Francis O, and Merriam J (2001) Beyond tides – Determination of core properties from superconducting gravimeter observations. *Journal of Geodetic Society of Japan* 47(1): 364–372.

Spratt RS (1982) Modelling the effects of the atmospheric pressure variations on gravity. *Geophysical Journal of the Royal Astronomical Society* 71: 173–186.

Suda N, Nawa K, and Fukao Y (1998) Earth's background free oscillations. *Science* 279: 2089–2091.

Sugihara M and Ishido T (2008) Geothermal reservoir monitoring with a combination of absolute and relative gravimetry. *Geophysics* 73(6): WA37–WA47.

Sugihara M and Nawa K (2012) Continuous gravity measurements for practical monitoring. In: *Proceedings of the Thirty-Seventh Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, California, January 30–February 1, 2012,* SGP-TR-194.

Sugihara M, Nawa K, Nishi Y, Ishido T, and Soma N (2013) Continuous gravity monitoring for CO_2 geo-sequestration. *Energy Procedia* 37: 4302–4309.

Sun H-P (1995) *Static Deformation and Gravity Changes at the Earth's Surface due to Atmospheric Pressure*. PhD Thesis, Obs. Roy. de Belgique, Bruxelles.

Sun H-P, Hsu H-T, Jentzsch G, and Xu J-Q (2002) Tidal gravity observations obtained with a superconducting gravimeter at Wuhan/China and its application to geodynamics. *Journal of Geodynamics* 33: 187–198.

Sun H-P, Jentzsch G, Xu J-Q, Hsu H-Z, Chen X-D, and Zhou J-C (2004) Earth's free core nutation determined using C032 superconducting gravimeter at station Wuhan/China. *Journal of Geodynamics* 38: 451–460.

Sun W, Miura S, Sato T, et al. (2010) Gravity measurements in southeastern Alaska reveal negative gravity rate of change caused by glacial isostatic adjustment. *Journal of Geophysical Research* 115: B12406. http://dx.doi.org/10.1029/2009JB007194.

Svenson S and Wahr J (2002) Estimated effects of the vertical structure of atmospheric mass on the time-variable geoid. *Journal of Geophysical Research* 107(B9): 2194. http://dx.doi.org/10.1029/2000JB000024.

Takemoto S, Fukuda Y, Higashi T, et al. (2002) Effect of groundwater changes on SG observations in Kyoto and Bandung. *Bulletin d'Informations des Marees Terrestres* 135: 10839–10844.

Tamura Y (1987) A harmonic development of the tide generating potential. *Bulletin d'Informations des Marees Terrestres* 99: 6813–6855.

Tamura Y (1993) Additional terms to the tidal harmonic tables. In: *Proceedings of the 12th International Symposium Earth Tides, Beijing 1993*, pp. 345–350. Beijing, New York: Science Press.

Tamura Y, Sato T, Ooe M, and Ishiguro M (1991) A procedure for tidal analysis with a Bayesian information criterion. *Geophysical Journal International* 104: 507–516.

Tanimoto T (2005) The oceanic excitation hypothesis for the continuous oscillations of the Earth. *Geophysical Journal International* 160: 276–288.

Tanaka Y, Okubo S, Machida M, Kimura I, and Kosuge T (2001) First detection of absolute gravity change caused by earthquake. *Geophysical Research letters* 28(15): 2979–2981.

Tanimoto T and Um J (1999) Cause of continuous oscillations of the Earth. *Journal of Geophysical Research* 104: 28723–28739.

Tanimoto T, Um J, Nishida K, and Kobayashi N (1998) Earth's continuous oscillations observed on seismically quiet days. *Geophysical Research Letters* 25: 1553–1556.

Timmen L, Gitlein O, Klemann V, and Wolf D (2012) Observing gravity change in the Fennoscandian uplift area with the Hanover absolute gravimeter. *Pure and Applied Geophysics* 169(8): 1331–1342.

Traer J, Gerstoft P, Bromirski PD, and Shearer PM (2012) Microseisms and hum from ocean surface gravity waves. *Journal of Geophysical Research* 117: B11307. http://dx.doi.org/10.1029/2012JB009550.

Tsubokawa T (1991) Absolute and superconducting gravimetry in Japan. In: *Proceedings of the 2nd Workshop: Non Tidal Gravity Changes Intercomparison Between Absolute and Superconducting Gravimeters, Conseil de l'Europe, Cahiers du Centre Europeen de Geodynamique et de Seismologie, Luxembourg, 11,* pp. 47–71.

Van Camp M (1995) Noise induced by the refrigerating device of a superconducting gravimeter in the seismological station of Membach (Belgium). *Bulletin d'Informations des Marees Terrestres* 123: 9302–9314.

Van Camp M (1999) Measuring seismic normal modes with the GWR C021 superconducting gravimeter. *Physics of the Earth and Planetary Interiors* 116: 81–92.

Van Camp M, Camelbeeck T, and Francis O (2002) Crustal motions across the Ardenne and the Roer Graben (north western Europe) using absolute gravity measurements. *Metrologia* 39: 503–508.

Van Camp M, de Viron O, and Warburton RJ (2013) Improving the determination of the gravity rate of change by combining superconducting with absolute gravimeter data. *Computers & Geosciences* 51(2013): 49–55.

Van Camp M, Stein J, Rapagnanai G, and Rivera L (2008) Connecting a Quanterra Data Logger Q330 on the GWR C021 Superconducting Gravimeter. *Seismological Research Letters* 79(6): 785–796.

Van Camp M, Vanclooster M, Crommen O, et al. (2006) Hydrogeological investigations at the Membach station, Belgium, and application to correct long periodic gravity variations. *Journal of Geophysical Research* 111: B10403. http://dx.doi.org/10.1029/2006JB004405.

Van Camp M and Vauterin P (2005) Tsoft: Graphical and interactive software for the analysis of time series and Earth tides. *Computers in Geosciences* 31(5): 631–640.

Van Camp M, Wenzel H-G, Schott P, Vauterin P, and Francis O (2000) Accurate transfer function determination for superconducting gravimeters. *Geophysical Research Letters* 27(1): 37–40.

Van Camp M, Williams DP, and Francis O (2005) Uncertainty of absolute gravity measurements. *Journal of Geophysical Research* 110: B05406. http://dx.doi.org/10.1029/2004JB003497.

Van Dam TM and Francis O (1998) 2 years of continuous measurement of tidal and nontidal variations of gravity in Boulder, Colorado. *Geophysical Research Letters* 25: 393–396.

Van Dam TM and Wahr JM (1987) Displacements of the Earth's surface due to atmospheric loading: Effects on gravity and baseline measurements. *Journal of Geophysical Research* 92(82): 1281–1286.

Van Dam T, Wahr J, Milly P, and Francis O (2001a) Gravity changes due to continental water storage. *Journal of Geodetic Society of Japan* 47(1): 249–254.

Van Dam T, Wahr J, Milly P, et al. (2001b) Crustal displacements due to continental water storage. *Geophysical Research Letters* 28(4): 651–654.

Van Ruymbeke M (1989) A calibration system for gravimeters using a sinusoidal acceleration resulting from a periodic movement. *Bulletin Géodésique* 63: 223–235.

Venedikov A and Viera R (2004) Guidebook for the practical use of the computer program VAV – Version 2003. *Bulletin d'Informations des Marees Terrestres* 121: 11037–11103.

Virtanen H (2001) Hydrological studies at the gravity station Metsahovi, Finland. *Journal of Geodetic Society of Japan* 47(1): 328–333.

Virtanen H and Mäkinen J (2003) The effect of the Baltic Sea level on gravity at the Metsähovi station. *Journal of Geodynamics* 35(4–5): 553–565.

Wahr JM (1985) Deformation induced by polar motion. *Journal of Geophysical Research* 90(B11): 9363–9368.

Wahr J, Han D, and Trupin A (1995) Predictions of vertical uplift caused by changing polar ice volumes on a viscoelastic Earth. *Geophysical Research Letters* 22: 977–980.

Wahr J, Molenaar M, and Bryan F (1998) Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research* 103(B12): 30,205–30,229.

Wang R (1997) Tidal response of the Earth. In: Helmut W, Zürn W, and Wenzel H-G (eds.) *Tidal Phenomena. Lecture Notes in Earth Sciences,* pp. 27–57. Berlin/Heidelberg: Springer.

Warburton RJ, Beaumont C, and Goodkind JM (1975) The effect of ocean tide loading on tides of the solid earth observed with the superconducting gravimeter. *Geophysical Journal of the Royal Astronomical Society* 43: 707–720.

Warburton RJ and Brinton EW (1995) Recent developments in GWR Instruments superconducting gravimeters. In: *Proceedings of the 2nd Workshop: Non-tidal gravity changes Intercomparison between absolute and superconducting gravimeters, Cahiers du Centre Européen de Géodynamique et de Séismologie, Luxembourg, 11,* pp. 3–56.

Warburton RJ, Brinton EW, Reineman R, and Richter B (2000) Remote operation of superconducting gravimeters. In: *Proceedings of the Workshop: High-Precision Gravity Measurements with Application to Geodynamics and Second GGP Workshop, Cahiers du Centre Européen de Géodynamique et de Séismologie, 17, Luxembourg,* pp. 125–136.

Warburton RJ and Goodkind JM (1977) The influence of barometric-pressure variations on gravity. *Geophysical Journal of the Royal Astronomical Society* 48: 281–292.

Warburton RJ and Goodkind JM (1978) Detailed gravity-tide spectrum between one and four cycles per day. *Geophysical Journal of the Royal Astronomical Society* 52: 117–136.

Weise A, Kroner C, Abe M, et al. (2009) Gravity field variations from superconducting gravimeters for GRACE validation. *Journal of Geodynamics* 48: 325–330.

Weise A, Kroner C, Abe M, et al. (2012) Tackling mass redistribution phenomena by time-dependent GRACE and terrestrial gravity observations. *Journal of Geodynamics* 59–60: 82–91.

Wenzel H-G (1994) Accurate instrumental phase lag determination for feedback gravimeters. *Bulletin d'Informations des Marees Terrestres* 118: 8735–8751.

Wenzel H-G (1995) Accurate instrumental phase lag determination for feedback gravimeters. In: Hsu HT (ed.) *Proceedings of the 12th International Symposium Earth Tides*, pp. 191–198. Beijing, New York: Science Press.

Wenzel H-G (1996a) Accuracy assessment for tidal potential catalogues. *Bulletin d'Informations des Marees Terrestres* 124: 9394–9416.

Wenzel H-G (1996b) The nanogal software: Earth tide processing package ETERNA 3.30. *Bulletin d'Informations des Marees Terrestres* 124: 9425–9439.

Widmer-Schnidrig R (2003) What can superconducting gravimeters contribute to normal mode seismology. *Bulletin of the Seismological Society of America* 93(3): 1370–1380.

Wilmes H, Boer A, Richter B, et al. (2006) A new data series observed with the remote superconducting gravimeter GWR R038 at the geodetic fundamental station TIGO in Concepción (Chile). *Journal of Geodynamics* 41: 5–13, The Superconducting Gravimeter as a Field Instrument Applied to Hydrology.

Wilson CR, Scanlon B, Sharp J, Longuevergne L, and Wu H (2012) Field test of the superconducting gravimeter as a hydrologic sensor, ground water. *National Ground Water Association* 50(3): 442–449.

Wziontek H, Cordoba B, Crossley D, et al. (2012) Stability and accuracy of relative scale factor estimates for Superconducting Gravimeters. *Geophysical Research Abstracts* 14, EGU2012-12642-1, EGU General Assembly 2012.

Xi Q-W (1987) A new complete development of the tide generating potential for the epoch J200.0. *Bulletin d'Informations des Marees Terrestres* 99: 6766–6781.

Xi Q-W (1989) The precision of the development of the tidal generating potential and some explanatory notes. *Bulletin d'Informations des Marees Terrestres* 105: 7396–7404.

Xu Y, Crossley D, and Herrmann R (2008) Amplitude and Q of ${}_0S_0$ from the Sumatra earthquake as recorded on superconducting gravimeters and seismometers. *Seismological Research Letters* 79(6): 797–805.

Xu J-Q, Sun H-P, and Ducarme B (2004a) A global experimental model for gravity tides of the Earth. *Journal of Geodynamics* 38: 293–306.

Xu J-Q, Sun H-P, and Yang X-F (2004b) A study of gravity variations caused by polar motion using superconducting gravimeter data from the GGP network. *Journal of Geodesy* 78: 201–209. http://dx.doi.org/10.1007/s00190-004-0386-1.

Zerbini S, Negusini M, Romagnoli C, Domenichini F, Richter B, and Simon D (2002) Multi-parameter continuous observations to detect ground deformation and to study environmental variability impacts. *Global and Planetary Changes* 34: 37–58.

Zerbini S, Richter B, Negusini M, et al. (2001) Height and gravity various by continuous GPS, gravity and environmental parameter observations in the southern Po Plain, near Bologna, Italy. *Earth and Planetary Science Letters* 192: 267–279.

This page intentionally left blank

3.05 Gravimetric Methods – Satellite Altimeter Measurements

DP Chambers, University of South Florida, St. Petersburg, FL, USA

© 2015 Elsevier B.V. All rights reserved.

3.05.1	**Introduction**	118
3.05.1.1	Altimeter Jargon	121
3.05.2	**Measuring Range**	122
3.05.2.1	Ionosphere Path Delay	122
3.05.2.2	Dry Troposphere Path Delay	123
3.05.2.3	Wet Troposphere Path Delay	123
3.05.2.4	Sea State Bias	125
3.05.2.5	Inverted Barometer Correction	128
3.05.3	**Satellite Orbit**	128
3.05.3.1	Orbit and Ground Track	128
3.05.3.2	Aliasing	130
3.05.3.3	Precise Orbit Determination	132
3.05.4	**Calibration and Validation**	134
3.05.4.1	Calibration at Single Site, Instrumented Platforms	134
3.05.4.2	Cal/Val from Comparison to Tide Gauges	135
3.05.4.3	The *T/P*: Jason-1 Calibration Phase	137
3.05.5	**Applications to Geophysics**	137
3.05.5.1	MSS, Gravity, and Bathymetry	137
3.05.5.2	SL Change	141
3.05.6	**Conclusions and Future Prospects**	143
3.05.6.1	Laser Altimetry	144
3.05.6.2	Delay-Doppler Altimeter	144
3.05.6.3	Wide Swath Altimeter	145
References		145

Glossary

Crossover point The geographic point where ascending and descending ground track passes cross.

Dynamic ocean topography DOT; also known as sea surface topography and dynamic topography. The elevation of the sea surface height above the geoid, caused by surface geostrophic currents. Variations range from ± 3 m.

Exact-repeat orbit A special type of orbit used for satellite altimetry. The ground track will repeat its pattern on the Earth's surface after a certain period of time.

Geoid The equipotential surface that corresponds to mean sea level of an ocean at rest that includes both the gravitational potential and the rotational potential.

Geoid height The height of the geoid above a reference ellipsoid.

Ground track Trace of the satellite nadir point on the surface of the Earth as the satellite orbits and the Earth rotates under it.

Ground track pass One segment of the ground track, moving either south to north (an ascending pass) or north to south (descending pass). Passes are ordered sequentially in time, starting from the beginning of the repeat period (see repeat period). Altimeter data are normally distributed as pass files.

Mean sea surface MSS; the time average of the sea surface height. This can be computed for a single satellite along a particular ground track pass, or as a grid from a combination of multiple satellites.

Nadir point Point where the vector from the satellite normal to the geoid intercepts the geoid.

Nodal day Time required for the Earth to make one rotation relative to the satellite orbit's line of nodes.

Reference ellipsoid First-order approximation of the geoid, which models the dominant equatorial bulge. Generally defined by an equatorial radius and a flattening coefficient. SSH is computed relative to a reference ellipsoid.

Repeat period Amount of time before a ground track repeats itself (see exact-repeat orbit). The stated repeat period is normally given as the closest integer number of days, instead of the exact time; for example, a 17.0505-day repeat is typically referred to as a 17-day repeat period.

Sea state bias SSB; a bias in altimeter range measurement that is related to surface waves and the surface wind state.

Sea surface height SSH; the elevation of the sea surface above a reference ellipsoid. Variations range from ± 100 m.

Sea surface height anomaly SSHA; deviation of SSH away from the time-average mean surface. Variations range from ± 40 cm in most of the ocean to ± 1 m for eddies.

Significant wave height SWH; average height of the 1/3 highest waves in the field of view.

Sun-synchronous orbit A special type of exact-repeat orbit where the orbital node rate equals the rate of revolution of the Earth around the Sun. This means that the angle between the orbital plane and the vector between the Earth and the Sun is constant.

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00063-4

Nomenclature

a	Semimajor axis of satellite orbit
c	Speed of light in a vacuum
E	Integrated density of free electrons along altimeter path
f	Radar altimeter frequency
H_{SAT}	Normal height of satellite above a reference ellipsoid
i	Inclination of satellite orbit
J_2	Second zonal harmonic of the Earth's gravity field
P	Satellite orbit period
P_R	Exact-repeat period
R	True satellite range (distance between satellite altimeter and ocean surface)
R_E	Earth's equatorial radius
R_{meas}	Raw range measured by an altimeter, including all biases and errors
R_P	Earth's polar radius
t	Time
ΔR	Corrections to range to remove biases and errors
ΔR_{dry}	Correction to range due to dry gases in atmosphere
ΔR_{iono}	Correction to range due to free electrons in ionosphere
ΔSSH_{IB}	Inverted barometer correction to SSH
Δx	Spacing between adjacent ascending and descending passes on a satellite ground track
ϕ	Latitude
λ	Longitude
μ_E	Gravitational constant of the Earth
τ_{tide}	Alias period of ocean tide
ω_E	Rotation rate of Earth-fixed reference frame about inertial reference frame
$\dot{\Omega}$	Rate at which ascending node of satellite orbit moves relative to inertial axis

Abbreviations

CM	Center of mass
DORIS	Doppler Orbitography and Radiopositioning Integrated by Satellite
DOT	Dynamic ocean topography
EMB	Electromagnetic bias
EOF	Empirical orthogonal function
ERS-1	Earth remote sensing satellite 1
GDR	Geophysical data record
GFO	Geosat Follow-On
GMSL	Global mean sea level
GPS	Global Positioning System
GRACE	Gravity Recovery and Climate Experiment
IB	Inverted barometer
MSS	Mean sea surface
PTR	Point target response
RMS	Root mean square
SLR	Satellite laser ranging
SSB	Sea state bias
SSH	Sea surface height above reference ellipsoid ($=H_{SAT}-R$)
SSHA	Sea surface height anomaly
SWH	Significant wave height
SWT	Science Working Team
T/P	TOPEX/Poseidon
TOPEX	TOPography EXperiment

3.05.1 Introduction

Satellite altimetry was first proposed in the 1960s as a way to obtain global, synoptic measurements of the sea surface. Previously, one could measure variations in the sea surface regularly in time from tide gauges, but only along coastlines. In the deep ocean, one could measure water properties like temperature, salinity, and currents from a ship, but it took many months to obtain the data along a single transect, and the transect was often not repeated for years. There were also many regions of the ocean that were never observed, due to the difficulty in reaching all the vast areas of the oceans with surface vessels.

Satellites, on the other hand, orbit the Earth approximately once every 90 min. Because the Earth rotates beneath them, the satellite's path traces out a ground track, so that after a matter of days, it has covered a significant percentage of the Earth's surface (Figure 1). Thus, a satellite is the perfect platform for an instrument to measure the Earth's sea surface height (SSH), or sea level (SL). The measurement can be made at regular intervals along the ground track – a 1s SSH interval is common, which corresponds to a spacing of about 6 km because of the satellite's motion. The spacing between adjacent ground tracks can also be adjusted by changing the period needed for the ground track to repeat (Section 3.05.3.1). Longer repeat periods lead to smaller distance between adjacent ground tracks at the equator. Thus, satellite orbit parameters can be adjusted to give a ground track that provides whatever repeat interval or ground track spacing is desired. In essence, an altimeter can be considered a global distribution of tide gauges that sample the SL regularly at intervals of several days instead of every minute or hour. One sacrifices the high temporal sampling of the tide gauge for a much denser spatial sampling.

Measuring SSH from space is conceptually simple. Imagine a satellite orbiting the Earth and that its position in some well-defined reference frame is known accurately at any time. A vector from the center of mass (CM) of the satellite normal to a mathematical model of the Earth's surface defines the location of the subsatellite point longitude (λ) and latitude (ϕ) at some time, t, as well as the height of the satellite above that surface along the

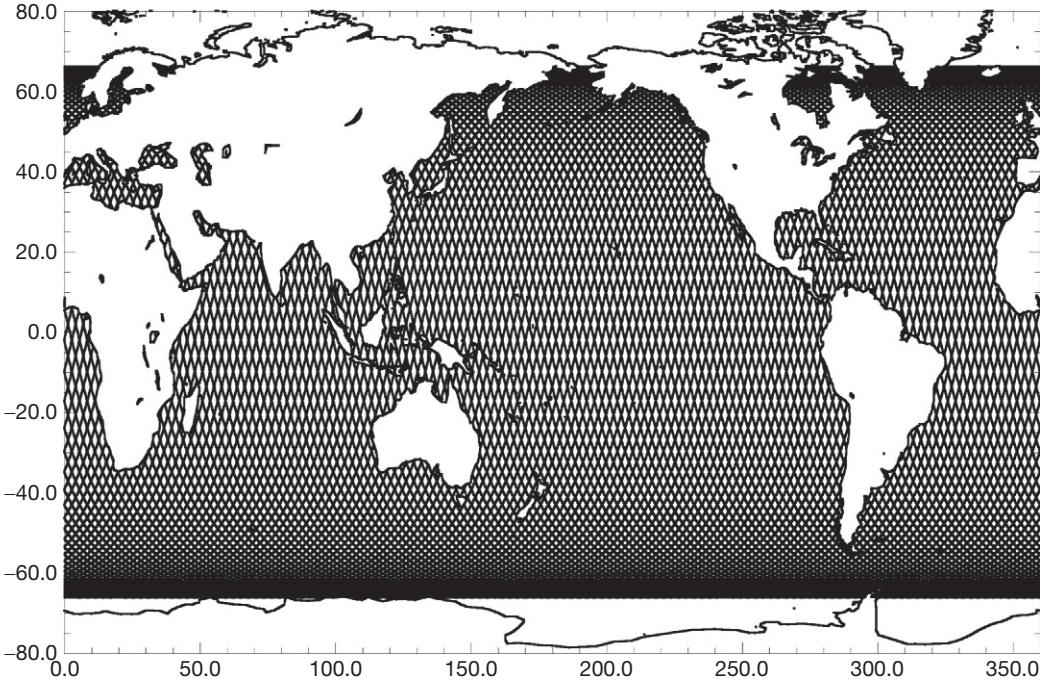


Figure 1 Ground track for TOPEX/Poseidon.

vector (H_{SAT}). For altimetry, this surface is generally modeled as an ellipsoid and not as a sphere to account for the Earth's oblateness at the equator. A commonly used ellipsoid is the one defined for the TOPEX/Poseidon (T/P) mission: equatorial radius = 6 378 136.5 m, flattening = 1/298.257. Then, if the satellite measures the distance between the radar antenna and the ocean surface below (R) via a ranging instrument, or altimeter, along this same nadir vector, then one can determine the SSH above the reference ellipsoid as a scalar quantity

$$\mathrm{SSH}(f, l, t) = H_{\mathrm{SAT}}(f, l, t) - R(f, l, t) \qquad [1]$$

In practice, the measurement is not so simple. The orbit of the satellite is not constant, but changes due to perturbations from drag, the Earth's nonuniform gravity, solar and Earth radiation pressure, and even the ocean tides. In order to determine H_{SAT} accurately, the satellite has to be continuously tracked, and its orbit estimated from the tracking data (e.g., Tapley et al., 2004a). Determining the range accurately is also not trivial. A radar altimeter consists of an antenna that both transmits a microwave pulse of known power and receives the partial power pulse reflected back from the ocean surface. To first order, the measured range is computed as

$$R_{\mathrm{meas}}(\phi, \lambda, t) = c\frac{\Delta t}{2} \qquad [2]$$

where Δt is the two-way travel time from initial transmission of the pulse to the arrival of the reflection and c is the speed of light in a vacuum. The travel time of the pulse needs to be known to be better than 10^{-10} s in order to ensure a range precision of 3 cm for a typical satellite altitude. Additionally, because the microwave pulse is traveling through the atmosphere and not a vacuum, corrections for attenuation need to be applied. The effective footprint of the pulse at the ocean surface is also of the order of 10 km, and radar energy is reflected differently from wave crests than from troughs, which will bias the range estimate. Thus, the true range is

$$R(\phi, \lambda, t) = R_{\mathrm{meas}}(\phi, \lambda, t) + \sum_i \Delta R(\phi, \lambda, t) \qquad [3]$$

where ΔR are a set of corrections to account for the path delay and the surface scattering effects. Ignoring any of these corrections can lead to errors of the order of tens of centimeters in range or more. Finally, the orbit height is defined to the CM of the satellite, while the range is measured relative to the phase center of the radar antenna, which is attached to the body of the satellite some distance from the CM. In order to recover SSH as accurately as possible, the vector from the antenna center to the CM needs to be known precisely, including any change caused by burning fuel to maintain the orbit.

Since the original development of the satellite altimeter concept, several missions have been flown, each progressively returning more accurate measurements of SSH as experience has been gained. The GEOS-3 mission was launched by the National Aeronautics and Space Administration (NASA) in April 1973, with an altimeter that measured range with a 25 cm precision for 1 s averages (Stanley, 1979). The initial orbit accuracy was about 5 m for the radial component (H_{SAT}). Although this meant that GEOS-3 could not observe the smallest variations in SSH, it demonstrated the capability and provided initial information on the approximately ± 100 m deviations of SSH relative to the ellipsoid. The satellite continued to operate for nearly 4 years. The Seasat mission was launched in July 1978 (also by NASA) with a significantly more precise altimeter (5 cm) as well as instruments to measure some of the atmosphere path delays (Lame and Born, 1982). The orbit determination had also improved, although

initial orbits (and SSH) still had errors of 1 m or more. Although it failed in 3 months due to a short circuit in the power system, Seasat demonstrated the usefulness of altimetry for applications beyond mapping the mean shape of the SSH.

Based on the hint of groundbreaking science suggested by Seasat, NASA began developing another altimeter mission. It was soon combined with a similar mission proposed by the French space agency, Centre Nationale d'Etudes Spatiales (CNES), and became known as T/P (the combination of the two original mission names). The goals of the T/P mission were to improve the altimeter range measurement (including both precision and corrections) as well as the orbit accuracy in order to reach a goal for SSH accuracy of 13 cm (Stewart et al., 1986).

The importance of having altimeter data for study before the launch of the new mission was recognized, especially with only 3 months of Seasat data. During the T/P development phase, the US Navy launched the Geosat altimeter mission on 12 March 1985, with a primary goal to map the marine geoid at very small wavelengths (McConathy and Kilgus, 1987). Several groups, especially those headed by Jim Mitchell and Bob Cheney, convinced the Navy to maneuver Geosat from its classified orbit into the unclassified Seasat orbit at the end of its primary mission (Born et al., 1987). It operated until December 1989 and provided a wealth of data for planning the new mission – most notably in efforts to improve orbit determination.

As noted, the goal of the T/P mission was to measure SSH to an accuracy of 13 cm root-mean-square (RMS), which was nearly an order of magnitude better than that accomplished up to that time. To meet this ambitious goal, the satellite was instrumented with more science equipment than any previous altimeter mission. This included two separate altimeters (one dual frequency), which could share the single antenna, a microwave radiometer to measure the delay due to water vapor along the path, and three independent tracking systems: a reflector for satellite laser ranging (SLR), a Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) tracking receiver, and an experimental Global Positioning System (GPS) receiver (Fu et al., 1994). The primary altimeter TOPography EXperiment (TOPEX) was based on the Seasat and Geosat altimeters with extensive modifications. These included adding another frequency to estimate ionosphere path delay directly (**Section 3.05.2.1**), increased precision (~2 cm RMS), and redundant circuits in order to prolong life (i.e., the primary TOPEX_A and a backup TOPEX_B altimeter). The secondary altimeter (Poseidon) was an experimental, solid-state altimeter that operated approximately 10% of the time in order to test the applicability of a low-power, low-mass altimeter for future missions.

One of the more important aspects of the T/P mission is often overlooked. A Science Working Team (SWT) was created 5 years before the anticipated launch date and tasked with the development of algorithms and models to meet the SSH accuracy goal. The team far exceeded the aim, and within 2 years of the launch in August 1992, SSH accuracy was estimated to be better than 4 cm RMS (Fu et al., 1994). One of the main reasons for the improved accuracy was better precise orbit determination (POD), due mainly to improvement in Earth gravity field models (Nerem et al., 1994a; Tapley et al., 1996).

At the same time NASA and CNES were developing the T/P mission, the European Space Agency (ESA) was developing its first Earth remote sensing (ERS-1) satellite. Although ERS-1 included a radar altimeter, the goals of the mission were significantly different than T/P. T/P was designed as a dedicated radar altimeter mission in order to obtain the most accurate measurement of SSH possible for scientific applications. The ERS-1 altimeter, on the other hand, was just one of many remote sensing instruments on the platform. This led to compromises in the mission due to requirements of the other instruments, mainly that the satellite was in a Sun-synchronous orbit at a relatively low altitude compared to T/P, so drag forces were much higher. In addition, one of the tracking systems (PRARE) failed shortly after launch in July 1991, so orbit accuracy was significantly poorer than what was available for T/P. The ERS-1 altimeter also operated at only a single frequency so could not directly measure the path delay in the ionosphere.

Although ERS-1 was launched over a year before T/P, data from its altimeter were not used widely until years later, unlike the data from the T/P mission, which many scientists embraced early. Part of this was the ease of use of the T/P data products, which were designed to be used by a wide variety of scientists without having to be an expert in radar altimetry. The ERS-1 data products, on the other hand, required more expertise to use. In addition, the range and orbits for the early T/P mission were significantly more accurate than those for ERS-1 and so were more useful at studying much smaller SSH signals. Finally, updates to important algorithms to improve the accuracy of T/P were incorporated directly into the processing stream and data products were reissued. ERS-1 algorithms were essentially fixed, and users had to collect algorithm changes from various published (and unpublished) sources and correct the data themselves (Scharroo, 2002).

Since the launches of ERS-1 and T/P, ESA has launched ERS-2 in April 1995 and Envisat in March 2002. ERS-2 carried essentially the same single-frequency altimeter as ERS-1, but the PRARE tracking system worked so that orbits were slightly better. Envisat carried both an improved two-frequency radar altimeter and a DORIS tracking receiver. Between the launches of ERS-1 and ERS-2, the US Navy launched a Geosat Follow-On (GFO) mission into the Seasat/Geosat ERM ground track in February 1998. The GFO altimeter was again only a single-frequency instrument, but unlike Geosat, the satellite had an SLR reflector array and a GPS receiver. Unfortunately, the GPS receiver failed early in the mission, and hence, orbit accuracy suffered. NASA and CNES launched a follow-on mission to T/P in December 2001: Jason-1. Jason-1 was designed to provide as accurate a SSH measurement as T/P in the same orbit, but using less power from a solid-state altimeter derived from the Poseidon instrument. A follow-on mission, Jason-2/Ocean Surface Topography Mission (OSTM), was launched in June 2008 in order to provide continuity of T/P and Jason-1 measurements. Jason-2/OSTM incorporates updated instruments derived from Jason-1 heritage, including an improved altimeter capable of tracking closer to shore, updated GPS hardware, and an advanced microwave radiometer (AMR) (Zaouche et al., 2010). The T/P mission ended in October 2005 due to a failure in the stability control mechanism. ESA lost contact with Envisat in April 2012 and abandoned attempts to reestablish contact, the following month. GFO was retired in November 2008 and deorbited. As of this writing, Jason-1 and Jason-2/OSTM are still operational.

It is not the intent of this chapter to provide every detail of how a radar altimeter works and how orbits are determined precisely or to document all the science that has been accomplished with every altimeter. Such detail is beyond the scope of a single chapter. For more details, the reader is pointed to the excellent book edited by Fu and Cazanave (2001), *Satellite Altimetry and Earth Sciences*, especially the first chapter by Chelton et al. (2001). Instead, we will summarize the important corrections necessary to obtain the most accurate altimeter range measurement and provide some background on types of altimeter orbits and how they are changed depending on the application, how the measurement is calibrated and validated, and how the data are used to study tides, SL change, and the shape of the geoid. More importantly, we will comment on various corrections that are needed for the raw T/P, Jason-1, and Jason-2/OSTM altimeter data that have been discussed extensively in Science Team meetings and mentioned in some literature, but have not been summarized completely. Although these corrections are applied to analysis products distributed by NASA, CNES, and other groups, as well as some releases of the geophysical data records (GDRs), there are still some versions of the GDRs that have none or only a few of these corrections. Thus, we believe it is important to summarize them in one location. We also tend to examine results of the T/P, Jason-1, and Jason-2/OSTM missions, as these data are widely used by scientists and the author is more familiar with them.

3.05.1.1 Altimeter Jargon

Before proceeding, it is worthwhile to note some jargon that is used in altimetry. Since these terms are often used in the current literature without definition, it is important to understand what they mean and will enable an easier discussion.

The *nadir point* of the orbiting satellite is the point where the vector from the satellite normal to the reference ellipsoid intercepts the ellipsoid. The nadir point will trace out a *ground track* on the surface of the Earth (Figure 1). Satellite altimeter missions are often in a special orbit that will repeat the ground track after a certain amount of time, the *repeat period*. This type of orbit is known as an *exact-repeat orbit*. A particular repeat period is often referred to as a *cycle*. Thus, Cycle 1 will refer to the first repeat period (or cycle) of the mission, while Cycle 100 will refer to the 100th repeat period. Different mission will have different repeat periods. For example, Geosat had a 17-day repeat period (exactly 17.0505 days), while T/P and Jason-1 have 10-day repeats (9.915625 days). See Section 3.05.3.1 for more details on how exact-repeat orbits are determined.

In satellite altimetry, the data are often distributed as files along a particular ground track, called a *pass*. The numbering of passes is determined by the time the pass crosses the equator relative to the time of the start of the repeat cycle. The first pass crosses the equator immediately after the start of the cycle, while the last pass crosses the equator just before the start of the next cycle. Passes are coincident in time, but not in location. Pass 2 will never be adjacent geographically to Pass 1, only adjacent in time. Passes are also often referenced by the direction the satellite is moving. For an *ascending pass*, the satellite is moving from south to north as it crosses the equatorial plane (Figure 2). A satellite is moving north to south on a *descending pass*. The geographic point where ascending and descending passes cross is referred to as a *crossover* point (Figure 2). One can determine crossovers from a single-satellite ground track (called a *single-satellite crossover*) or between ground tracks of different altimeter satellites (called a *dual-satellite crossover*).

Note that ascending passes do not always slope south to north from west to east, while descending passes slope north to south from west to east. This pattern is only valid for satellites with an inclination less than 90°. If the inclination is greater than 90°, then ascending passes slope south to north from east to west and similarly for descending passes (Figure 3).

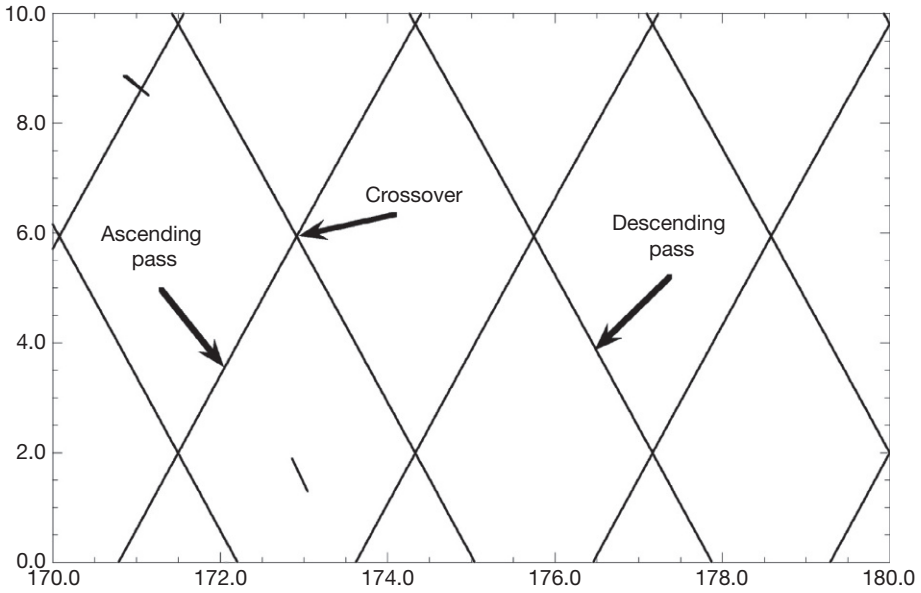


Figure 2 Crossover point for T/P, showing ascending and descending passes.

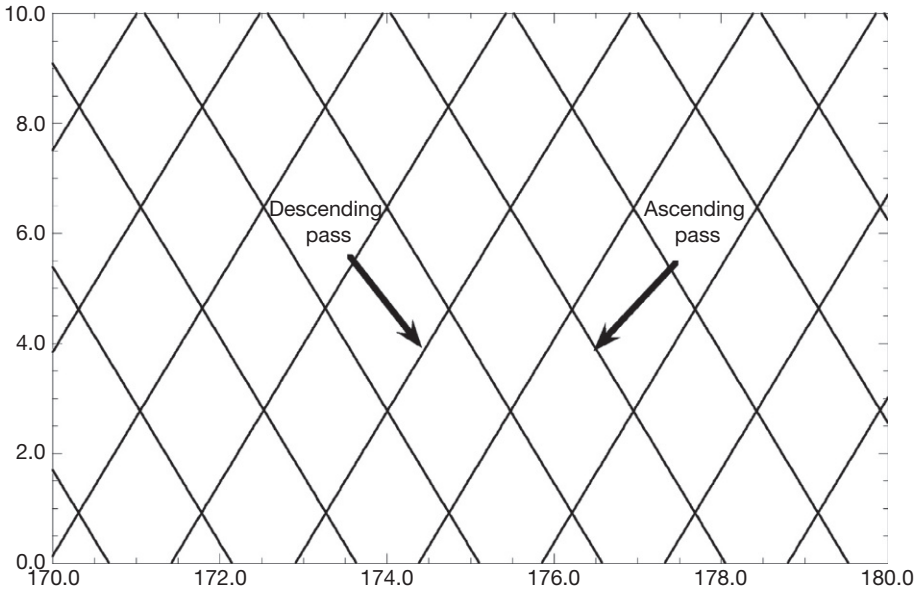


Figure 3 Crossover point for Geosat, showing ascending and descending passes.

3.05.2 Measuring Range

A radar altimeter consists of a radio transmitter that emits very sharp pulses of microwave energy, a receiver to record the pulse after it has been reflected from the ocean surface, and a clock to record the time between transmission and reception. Several thousand pulses are transmitted each second. Many of these pulses must be averaged together to obtain a smooth estimate of the pulse shape because the instantaneous pulse is the superposition of the scatter from many wave facets on the sea surface. In reality, the processing of the altimeter signal is far more complicated than this, as the pulse shape is changed because of the reflection and becomes more of a stretched waveform, and this is what the time of transit is estimated from. The precision of the range will also be dependent on how many waveforms are averaged before determining the transit time. For this discussion, we will assume that a two-way travel time has been determined from the waveforms at a 1s (1 Hz) spacing, which is typical for radar altimeters. For more details on how the transit time is actually calculated, the reader is directed to Chelton et al. (2001).

As discussed before (eqn [2]), the raw range measurement is then calculated from the two-way transit time assuming that the speed of light is constant. However, this is only true in a vacuum, and moving through a medium like the atmosphere will slow down the electromagnetic wave. Thus, the raw range measurement will be too long, with the error depending on the properties of the atmosphere. It must be corrected in order to determine an accurate range. In altimetry, the corrections are given in terms of the negative path delay that needs to be added to the range in order to determine the correct value (eqn [3]).

3.05.2.1 Ionosphere Path Delay

Free electrons in the Earth's ionosphere will refract the radar pulse and delay the path, ΔR_{iono}, by an amount proportional to the density of free electrons along the path (E) and inversely proportional to the square of the frequency (f) of the microwave pulse

$$\Delta R_{\text{iono}}(f) = -\frac{40.25E}{f^2} \qquad [4]$$

where E is the integrated density of free electrons (electrons m^{-2}) along the path and f is the radar frequency in hertz. The units of the path delay will be in m and the negative sign is following the convention set by T/P: that all path delay corrections are added to the raw range. Thus, the effect of the correction is to shorten the raw range measurement.

The electron count varies over the day (more free electrons in the day), over the seasons (more free electrons in the winter), over the 11-year solar cycle (more during the solar maximum), and meridionally (maximum at the magnetic equator). Typical values of E range from 10^{16} to 10^{18} electrons m^{-2}, which implies, for a 13 GHz altimeter, errors in the range of 2 mm to 20 cm. If E is known precisely, this range error could be removed completely. However, before T/P, altimeters relied on models of the electron density to determine the path delay.

One way to determine the path correction directly is to operate a two-frequency altimeter. The path delay will be different for each frequency band, as will the raw range measurement, but the corrected range measurements should be identical, that is,

$$\begin{aligned} R &= R_1 + \Delta R(f_1) \\ R &= R_2 + \Delta R(f_2) \end{aligned} \qquad [5]$$

where the subscript 1 denotes the primary frequency for science and 2 denotes the secondary frequency. If these two equations are differenced, and the path delay equation (eqn [4]) is substituted, then

$$\begin{aligned} R_1 - R_2 &= -40.25E\left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right) \\ R_1 - R_2 &= -40.25E\left(\frac{f_1^2 - f_2^2}{f_1^2 f_2^2}\right) \\ \Delta R(f_1) &= \frac{f_2^2}{f_1^2 - f_2^2}(R_1 - R_2) \end{aligned} \qquad [6]$$

Thus, given the measured ranges at the two known frequency bands, one can obtain the ionosphere path delay correction directly. Generally, the only correction applied to the raw ranges before combining is the sea state bias (SSB) (**Section 3.05.2.4**), as this is the only other frequency-dependent range correction.

For T/P, Jason-1, and Jason-2/OSTM, the primary science frequency is in the Ku-band (13.6 GHz), while the secondary band is in the C-band (5.3 GHz). The reason for this is that the Ku-band is affected less by the ionosphere than the C-band (eqn [4]) for a particular electron density and the instrumental errors for the Ku-band are less (Chelton et al., 2001). The only other dual-frequency altimeter that has operated is Envisat, using Ku-band (13.575 GHz) and S-band (3.2 GHz) frequencies. The estimated error in the ionosphere delay correction derived from the dual-frequency TOPEX has been estimated to be less than 1 cm RMS (Imel, 1994; Zlotnicki, 1994). The Jason-1 correction agrees with that of TOPEX to 1.3 cm RMS, indicating an accuracy comparable to TOPEX or even slightly better (Chambers et al., 2003a). A similar analysis between Jason-1 and Jason-2 ionosphere path delays shows that except for a small bias, the corrections agree to be better than 1 mm RMS (Ablain et al., 2010).

Single-frequency altimeters rely on models of the ionosphere electron density and eqn [4] in order to estimate a correction. Original models for Seasat and Geosat had errors in the electron density as large as 50% (Cheney et al., 1991), but subsequent models (e.g., Bilitza, 1997; Daniell et al., 1995) have reduced the errors to 20% or less. However, this is still equivalent to path delay errors as large as 3 cm RMS, and these models will never match the accuracy of the correction from a dual-frequency altimeter.

Although the Poseidon altimeter was also a single-frequency altimeter and cannot measure the electron content directly, T/P carried a DORIS tracking receiver for POD (see **Section 3.05.3.3**). DORIS beacons at approximately 50 locations around the world transmit at two frequencies (0.4 and 2.0 GHz) in order to correct for the ionosphere path delay in the measurement between the satellite and the station. This allowed for the calculation of electron content along a path between T/P and the DORIS ground stations. In order to be used by an altimeter, though, the slant measurement had to be interpolated to the radar beam path, usually by using an empirical model (Fleury et al., 1991). Although the results comparing the DORIS ionosphere corrections with those obtained from TOPEX for several periods when the two operated within the same 10 day cycle indicated the DORIS model was better than the empirical models, the global RMS error is still about 2 cm (Imel, 1994; Zlotnicki, 1994). More troubling, Zlotnicki (1994) showed that the signal was correlated over very large distances, indicating long-wavelength errors.

Another method for determining total electron content (TEC) is from analyzing the two frequencies transmitted by GPS satellites and detected by ground receivers (e.g., Mannucci et al., 1999). As with the determination from DORIS, this technique will only determine TEC accurately for vectors from ground stations to the GPS satellites. To produce global maps, the data have to be interpolated. Comparisons with ionosphere path delay derived from these TEC maps with those derived from the dual-frequency TOPEX altimeter indicate that while differences near GPS ground stations are small (less than 1 cm RMS), the difference away from ground stations can be up to ten times larger (Imel, 1994; Schreiner et al., 1997). Thus, although models of global electron density are improving, none can meet the accuracy of 1 cm RMS that can be achieved with a dual-frequency altimeter.

3.05.2.2 Dry Troposphere Path Delay

The radar pulse is also slowed by gases in the Earth's troposphere. The amount of 'dry' gases is nearly constant and produces height errors with a magnitude of about 2 m. However, this path delay does not change appreciably over time. The path delay (ΔR_{dry}) can be related to the sea level pressure (SLP) (P_0) and latitude (e.g., Chelton et al., 2001) as

$$\Delta R_{\text{dry}} \approx -0.2277 P_0 (1 + 0.0026 \cos 2\phi) \qquad [7]$$

which has units of cm if P_0 is in mbar.

The only practical way to obtain global SLP data is from numerical weather prediction models that assimilate available SLP measurements, such as those from the National Center for Environmental Prediction (NCEP) or the European Center for Medium-Range Weather Forecasting (ECMWF). The data from the ECMWF model are used in most modern altimeters such as T/P, Jason-1, Jason-2/OSTM, and Envisat, as its SLP data are widely considered more accurate than those from NCEP (Trenberth and Olson, 1988). Although SLP pressures are typically high (about 950–1050 mbar), meaning the total range correction is about 2.3 m, the correction is not very sensitive to errors in SLP. For example, a 1 mbar error in SLP is equivalent to a 2 mm error in the path delay.

An approximate measure of the SLP error is a comparison of the output of the ECMWF and NCEP models, although this will not account for similar errors that cancel. Typical RMS differences are 2–3 mbar equatorward of 40° latitude, and 4–7 mbar poleward of 40° latitude (Chelton et al., 2001), which suggests errors in the path delay of between 5 mm RMS in the tropics and sometimes larger than 1 cm at higher latitudes.

3.05.2.3 Wet Troposphere Path Delay

The path delay caused by both water vapor and cloud liquid water can be determined using a passive microwave radiometer. As mentioned earlier, Geosat did not carry a microwave radiometer so had to rely on data from models or other satellite radiometers. The wet troposphere correction for Geosat was originally based on the Fleet Numerical Oceanography Center (FNOC) operational analysis of humidity. This was shown to significantly degrade Geosat ranges in the tropics where the FNOC model consistently underestimated water vapor (Emery et al., 1990; Zimbelman and Busalacchi, 1990). Later updates to the data used a correction based on interpolating water vapor measurements determined from other satellites (SSMI and TOVS) to the Geosat ground track (Cheney et al., 1991), but the error was still of the order of 4 cm RMS (Shum et al., 1993).

A two-channel radiometer, operating at two frequencies near 20 GHz, is the minimum required to determine the path delay. The two frequencies are necessary to remove factors affecting the microwave brightness temperatures other than water vapor.

However, it has been shown that operating a downward-looking satellite radiometer with a third frequency allows one to correct for effects of wind-induced changes to sea-water emissivity (e.g., Keihm et al., 1995). Both the TOPEX microwave radiometer (TMR) and the Jason-1 microwave radiometer (JMR) operate at three frequencies: 18.0, 21.0, and 37.0 GHz. The AMR flown on Jason-2/OSTM operates at 18.7, 23.8, and 34.0 GHz. Overall, the estimated accuracy of the T/P path delays is estimated to be about 1 cm (Ruf et al., 1994) with estimates for Jason-1 being even lower (e.g., Brown et al., 2004). Keihm et al. (1995) estimated that the error in wet troposphere path delay would be about 1–2 cm higher if the 37 GHz channel were not used, which is consistent with the finding for the ERS-1 and ERS-2 missions, which use a dual-frequency radiometer at 23.8 and 36.5 GHz (Eymard et al., 1996).

One problem with microwave radiometers is that they have historically suffered from calibration drifts. Although there was considerable work undertaken to minimize such drifts in the TMR, a drift became apparent in the 18.0 GHz channel over the course of the mission. This effect was a nearly linear drift in path delay of approximately 1.5 mm year^{-1} over the first 4 year of the mission (Keihm et al., 2000; Scharroo et al., 2004) and was first observed in the comparison of SSH with tide gauges (Chambers et al., 1998a; Mitchum, 1998; see also **Section 3.05.4.2**). The sense of the error was to make the measured SSH lower than it actually was. More recent analysis suggests that the drift lasted for about 6.5 years before leveling off (Scharroo et al., 2004). We should note that not every T/P GDR or data product that is available to the public has been corrected for this error. Therefore, the user needs to verify if this correction has been applied to T/P data or correct for it (e.g., Ruf, 2002; Scharroo et al., 2004).

There was also a smaller change in the TMR wet path delay that varied in time. It was not noticed until the TMR and JMR path delays were compared early in the T/P–Jason calibration phase (Chambers et al., 2003b). It was observed that the two path delays changed abruptly by nearly a centimeter at several periods during the 210-day period before returning to normal. It was quickly established that the error was in the TMR measurement and in fact had existed for the entire T/P mission (Chambers et al., 2003a; Zlotnicki and Desai, 2004). The timing was eventually linked to transition from one yaw state to another and has also been shown to exist in the JMR path delays, but to a smaller extent (Brown et al., 2004; Desai and Haines, 2004; Obligas et al., 2004; Zlotnicki and Desai, 2004). The cause for the correlation with the transition from one yaw state to another was linked to higher than normal temperatures on the MR feedhorn that had not been simulated when determining the MR calibration coefficients (Brown et al., 2004). New calibration coefficients have been determined to remove this effect in the JMR data (Brown et al., 2004) and were used to generate the first release Jason-1 GDRs. A similar calibration has not been applied to the TMR data in the T/P GDRs.

In addition to the yaw-mode path delay biases, the JMR has also suffered several additional biases in the path delay. The first occurred in November 2002, when the path delay correction became biased by about 3–4 mm (Desai and Haines, 2004; Zlotnicki and Desai, 2004). Approximately a year later, it became biased again by a further 7–8 mm in the same direction. The changes can be seen clearly in **Figure 4**, which shows the mean difference with path delay estimated from ECMWF water vapor averaged for each Jason-1 repeat cycle. The cause has been linked to thermal shock to the instrument, and the error will bias the SSH measurement globally by the same amount as the path delay bias. Although original release Jason-1 GDRs were affected by this problem, subsequent reprocessed GDRs have been corrected using adjusted calibration coefficients.

The ERS and Envisat radiometers are also not immune to changing biases in the path delay. The ERS-2 radiometer suffered a far larger change in the 23.8 GHz channel than has been observed in either the TMR or JMR channels, resulting in a 10 K change in brightness temperature after 16 June 1996 (Scharroo et al., 2004). The Envisat radiometer also appeared to have undergone a 1 K change in brightness temperature for both channels starting in 2003 (Scharroo et al., 2004), and a new algorithm to correct for these has been developed (Obligis et al., 2007). Corrections for these drifts have not necessarily

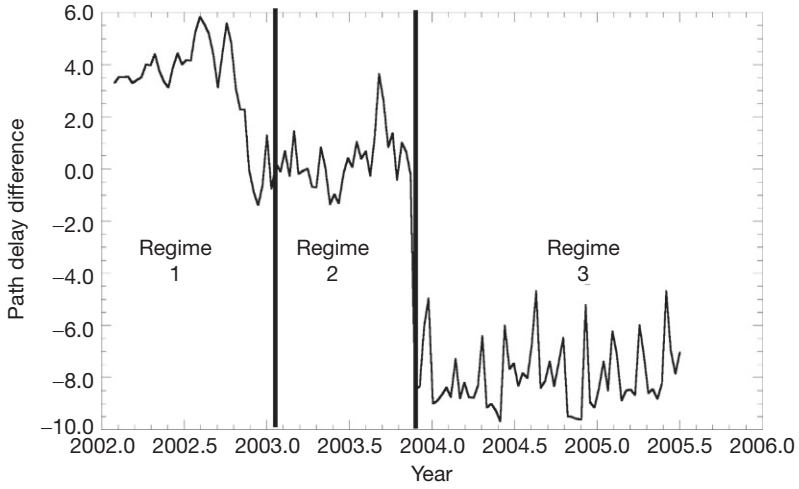


Figure 4 Path delay from JMR minus path delay estimated from ECMWF model stored on GDRs, averaged globally for each repeat cycle. Note that path delays are stored as negative values on GDRs but that here we have differenced the positive magnitudes.

been applied to all released data records, so users are cautioned to read the data documentation carefully.

In order to use altimetry for climate studies related to SL change (**Section 3.05.5.2**), the SSH measurements must be relatively stable. Because of the problems that have been identified in the radiometers on all altimeter missions, an autonomous radiometer calibration system has been developed at JPL to calibrate the AMR data before GDRs are processed. It is based on comparing measured brightness temperatures with predicted temperatures from cold ocean regions (Brown et al., 2009; Ruf, 2002; Scharroo et al., 2004), or from warm Earth areas that act as a pseudo blackbody emitter, like the Amazon (Brown et al., 2009). Studies have found that these calibrations are accurate to the 1 K level, once slight climate variations in the warm regions are accounted for (Brown et al., 2009). Although these same types of calibrations were used previously, they were typically done only once or twice a year, were not autonomous, and were not fed into the processing stream. The new system for AMR resolves these issues so that new calibrations are applied as needed, and changes can be detected quickly. This is especially important, as the 34.0 GHz channel of AMR has drifted by 6 K since launch, and the other channels have also shown abrupt changes since 2010. After applying these corrections, the AMR is reasonably stable, although there have been several jumps equivalent to 3–4 mm of path delay based on comparison with path delay derived from near-coincident GPS data (Sibthorpe et al., 2011) and there also still appears to be a small residual drift in the 23.8 GHz channel of -0.3 K year^{-1} (Cao et al., 2011). This latter estimate was based on comparing AMR brightness temperatures with those from a more stable radiometer, the advanced microwave sounding unit (AMSU). Instead of relying on Earth-based emitters for calibration, the AMSU calibrates against a cold target in deep space and a blackbody radiator carried onboard, with a known temperature. Such internal calibration results in a much better stability than external calibration (Cao et al., 2011) and is being investigated for use on future altimeter radiometers.

3.05.2.4 Sea State Bias

The range is also biased due to surface waves in the footprint of the radar pulse. The bias is composed of three components. The first, known as the electromagnetic bias (EMB), arises because wave troughs backscatter more power per unit area than wave crests. This creates a bias in the range toward the troughs and away from the desired measurement of mean SL. The second effect is known as the skewness and is caused by the fact that the onboard tracker determines the two-way travel time from the half power point of the waveform that corresponds to the median scattering surface and not the mean scattering surface (Chelton et al., 2001). The third component relates to how the wave forms are tracked and range estimated so is referred to as the tracker bias. Although these biases are distinctly different and related to different instrumental effects, they are difficult to separate completely as they are all related to the size of waves in the footprint. Several studies have attempted to determine a theoretical value for the EMB portion alone, based on assumptions about the wave number spectrum (Glazman et al., 1996; Rodriguez et al., 1992; Srokosz, 1986). However, as more research has been conducted, it has become more and more apparent that the EMB also tends to be affected by the radar frequency as well as wind speed (Melville et al., 1991; Walsh et al., 1991) and so varies from one instrument to another. Because a part of the skewness and tracker biases are also linearly dependent on significant wave height (SWH) (e.g., Hayne and Hancock, 1982; Srokosz, 1986), any empirical estimate of the EMB from altimeter data will include the portion of the skewness and tracker bias that is related to SWH. Because of this, one usually considers the combined effect as the SSB. Although theoretical studies of the EMB portion continue (Elfouhaily et al., 1999, 2000), they still cannot adequately explain the full sea state bias in altimetry. Therefore, in practice, the combined SSB is estimated empirically from each altimeter data so as to reduce the variance of the SSH data from that altimeter.

SSB is expressed as a linear function of SWH, which is the average height of the 1/3 highest waves in the field of view

$$\mathrm{SSB} = -b\mathrm{SWH} \qquad [8]$$

where b is determined empirically. Wave conditions cause changes in the slope of the leading edge of the radar altimeter waveform. High slopes indicate small waves, while shallow slopes indicate large waves, very nearly with a linear relationship. The power of the returned waveform is also related to the roughness of the surface, caused by the wind speed. Chelton et al. (2001) described completely the history and process for determining SWH and wind speed from the waveforms. In this discussion, we will assume that a measure of SWH and wind speed is returned at the same time as altimeter range and SSH.

Determining an empirical estimate of SSB is relatively straightforward. First, one determines SSH with all corrections except SSB at a particular location along a ground track (x) and at particular time (t), which to first order should be the true SSH ($\mathrm{SSH_{true}}$) corrupted by the SSB (bSWH), or

$$\mathrm{SSH}(x,t) = \mathrm{SSH_{true}}(x,t) + b\mathrm{SWH}(x,t) \qquad [9]$$

If the SSH was measured in the same location again some Δt later, then the residual (ΔSSH) will be

$$\begin{aligned}\mathrm{DSSH} &= \mathrm{SSH}(x,t+\mathrm{D}t) - \mathrm{SSH}(x,t)\\ &= b(\mathrm{SWH}(x,t+\mathrm{D}t) - \mathrm{SWH}(x,t)) + \mathrm{e}\end{aligned} \qquad [10]$$

assuming that the difference in the true SSH is small from one time to another. In reality, this is not necessarily true, but with a large enough number of observations, it should be in a mean sense, provided there are no biases. The factor ϵ accounts for any errors in the data and unmodeled variations. Thus, one seeks to find the value of b to minimize the sum of ϵ^2 over all observations:

$$\sum_{x,t} \epsilon^2 = \sum_{x,t} [\Delta\mathrm{SSH} - b\Delta\mathrm{SWH}]^2 \qquad [11]$$

Differences in the values estimated for b arise from how SSH is corrected, particularly if orbit error corrections are applied (Chelton et al., 2001), how many residuals are used, and what types of residuals are used. For example, early estimates for Seasat and Geosat used residuals based on differencing SSH along ground track passes from adjacent repeat cycles. One problem with this method is that there were fairly large geoid errors that

did not completely cancel out because the observations were not made at exactly at the same location. Altimeter crossovers, which are derived from interpolating the SSH to the point where ascending and descending ground tracks cross, not only do a better job of removing the geoid error but also result in far fewer points – there are a few thousand crossovers for a 10 day T/P cycle compared with nearly 400 000 collinear points. However, as Gaspar et al. (1994) pointed out, collinear points are usually correlated over fairly large distances and so are not completely independent, while crossovers are. Also, the time gap between crossovers is always smaller than the repeat cycle. For T/P, it is on average 3.5 days. This will reduce effects of the unmodeled SL variations that are ignored in the estimation process.

Another factor that affects the result is how b is modeled. The earliest models assumed b was constant everywhere over the ocean; that is, the SSB was just a percentage of SWH. However, Witter and Chelton (1991) found evidence in the Geosat data that b was not uniform and Fu and Glazman (1991) proposed another model that was a nonlinear function of both SWH and wind speed. This led to the model adopted by the T/P project before launch, which was empirically derived from aircraft measurements (Hevizi et al., 1993):

$$b = a_0 + a_1 U + a_2 U^2 \qquad [12]$$

where U is the wind speed and the parameters a_0, a_1, and a_2 depend on the frequency of the altimeter. The prelaunch algorithm was evaluated in several studies (Chelton, 1994; Gaspar et al., 1994; Rodriguez and Martin, 1994). It was found that a function of the form

$$b = a_0 + a_1 U + a_2 U^2 + a_3 \mathrm{SWH} \qquad [13]$$

resulted in approximately the same variance of residuals within $\pm 40^\circ$ as the original parameterization (eqn [12]) but significantly lowered the variance poleward of 40°, where wave heights and wind speeds tend to be higher. This 4-parameter model (eqn [13]) of Gaspar et al. (1994) was adopted for reprocessing of the T/P data.

Ideally, the SSB would not assume an a priori parametric model, but would be determined as a map of SSB versus SWH and U that could be used as a lookup table. Such nonparametric models were originally studied by Witter and Chelton (1991), Chelton (1994), and Rodriguez and Martin (1994), but did not show significant improvement over a parametric model. In preparation for the Jason-1 mission, Gaspar and Florens (1998) proposed a new method to calculate a nonparametric SSB model and demonstrated that their model further reduced the variance in midlatitude and high latitude by about 1 cm^2.

Errors in the SSB are difficult to quantify. Chelton (1994) found that estimating a parametric model with different data sets amounted to differences of about 1% of SWH, which he took as one estimate of the error. One thing to note is that the error will have a time-variable component as well as a bias component, as there will always be some waves in the ocean on average. This bias is also not constant, since the mean wave state varies from point to point. Assuming the 1% of SWH value is a reasonable upper bound on the error, then the maximum potential bias and variable errors are plotted in Figure 5. The time-variable portion of the error is probably smaller than 2 cm RMS except for areas with the highest waves. More troubling is the potential of systematic bias errors as large as 3–4 cm. This is especially a problem for studies of mean SL change, which depend on linking data from two different altimeters. However, for studies that remove a local mean of the SSH in order to observe the time-varying SSH, any bias in the SSB model will be removed, provided means are computed for each set of altimeter data.

Several investigators on the T/P SWT noticed significant changes in the observed SWH starting in mid-1997. It was confirmed via comparison with buoys that the TOPEX SWH measurement was in error. The size of the change in SWH was about 50 cm over 2 years and would have caused a change in SSB of about 1 cm if there had not been a negating factor. Instrument engineers found that the source of the problem was a noticeable degradation in the TOPEX altimeter point-target response (PTR), which would affect not only the SWH but also the range measurement (Fu, 1998), and did so in such a way as to nearly cancel the SSB effect. However, the PTR degradation was only expected to increase, and since there was a redundant altimeter onboard that had never operated (TOPEX_B), it was recommended that the original primary science altimeter (TOPEX_A) be turned off and TOPEX_B be turned on for at least a few cycles to see if this corrected the problem. This was done in February 1999. All tests indicated TOPEX_B performed as well as early TOPEX_A data, so the decision was made to keep TOPEX_B on as the primary science instrument for the remainder of the mission. The last T/P 10-day repeat cycle to use TOPEX_A was Cycle 235, ending on 9 February 1999. TOPEX_B was turned on at the beginning of Cycle 236 on the same day.

Other than a few initial calibrations based on the first few months of data, no significant changes were made to the data algorithms, in particular the SSB algorithm, or to the tracker. Shortly after the switch, it was noticed that there was a slight offset in the TOPEX_B data relative to tide gauge data (see **Section 3.05.4.2** for discussion) but the nature of the error was not clear until Jason-1 was launched and investigators began comparing Jason-1 data with TOPEX_B data. Chambers et al. (2003b) noticed that there was a residual relationship between the TOPEX_B and Jason-1 data as a function of SWH, even though coefficients to a SSB model had been adjusted for the new Jason-1 altimeter. They found that the relationship was reduced only if the SSB model for TOPEX_B was also adjusted, suggesting an error in the model. Upon further investigation, they demonstrated that the offset between TOPEX_B and TOPEX_A SSH that had been observed in the tide gauge calibration was removed when separate SSB models were used for TOPEX_A and TOPEX_B. An important consequence of the error in the TOPEX SSB models was that the step function was not globally uniform (**Figure 6**), as suggested by the tide gauge calibration. The change varied from 5 to 6 mm in the tropics (where most of the tide gauges for the calibration were located) to more than 2 cm in the far Southern Ocean. The effect of the erroneous SSB for TOPEX_B on the determination of global mean sea level (GMSL) rate was the decrease in the real measurement by more than 1 mm $year^{-1}$ (Chambers et al., 2003b).

This highlights why empirical estimation of altimeter SSB models is still so important. Theoretically, the TOPEX_A and TOPEX_B instruments were identical, with the exact same operating frequencies. However, there were clearly differences

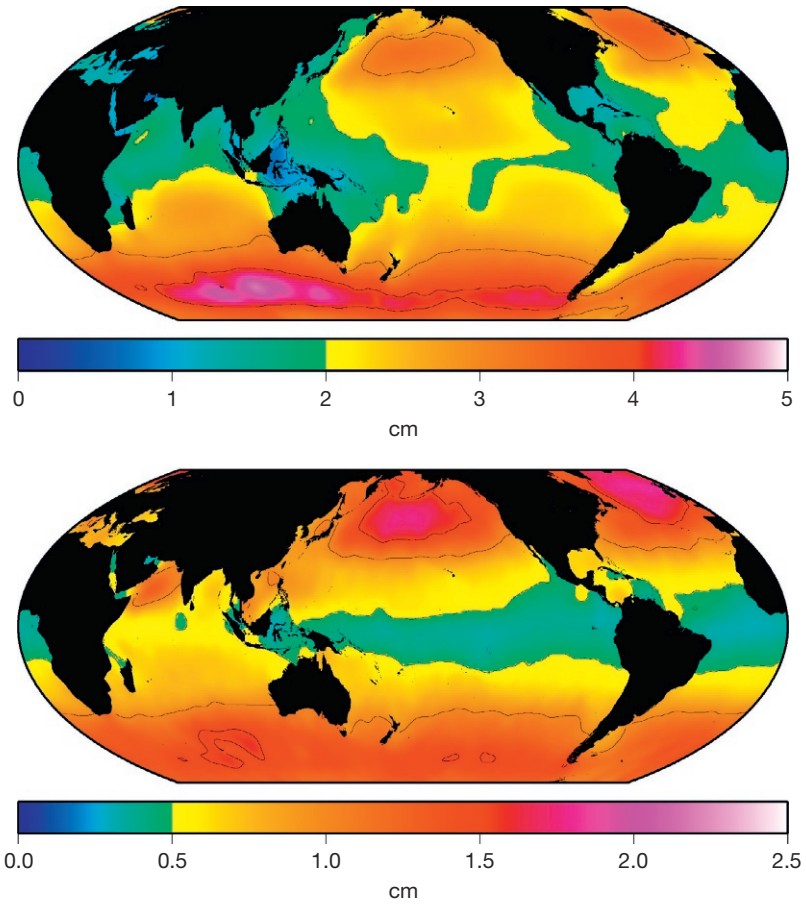


Figure 5 Estimates of possible errors in SSB. Mean error (top) is computed assuming error is 1% of mean SWH (Chelton, 1994). The variable error (bottom) is computed assuming error is 1% of the RMS of SWH. Note that the scale of the variable portion is half the size of the mean portion.

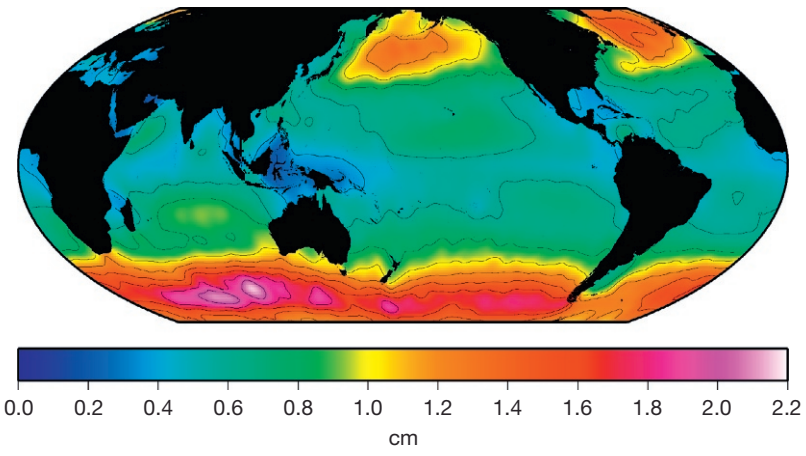


Figure 6 Mean bias correction to TOPEX_B data by switching from old SSB model to new model.

in the tracker bias, and this leaked into the SSB. Work on nonparametric models has continued, and recently, models have been developed for TOPEX_A, TOPEX_B, Jason-1, and other altimeters (Gaspar et al., 2002; Labroue et al., 2004; Tran et al., 2010). There has been continuing debate as to whether crossover points or collinear differences are best to use for estimating SSB. Crossovers generally have a shorter time separation, thus better representing sea state changes and not SSH changes. However, there are significantly larger numbers of collinear differences, but they are always separated in time by 10 days for T/P, Jason-1, and Jason-2 ground tracks. Labroue et al (2006) argued that the increased number of observations

using the collinear method made up for the longer time gap and this was the data set used for the model adapted in Jason-1 GDR reprocessing. More recently, however, Tran et al. (2010) found significant biases in SSB at high wave heights between different methods and proposed a modified approach that is based on the average of two different SSB models, one from 10 day collinear differences and the other based on SL anomalies relative to a mean sea surface (MSS). This later method is based on an approach first introduced by Vandemark et al. (2002) and later updated (Tran et al., 2006). This modified collinear method has been used to compute SSB for Jason-2, Jason-1, and T/P, although the SSB models for Jason-1 and T/P have to be applied to published GDRs.

Unfortunately, SSB models likely still have errors of up to several centimeters at for high wave heights and also can have biases of several centimeters between nonparametric and parametric models based on the same data or between using collinear and crossover approaches on the same data. Hausman and Zlotnicki (2010) recently tested various approaches to estimating SSB on output from an ocean model. Since the model has no inherent SSB, estimation of SSB using the output should result in zero values. This was not the case, however. Using collinear differences from consecutive cycles, the recovered SSB model had a global bias of about −1 cm and a standard deviation of 5 cm. Even higher biases and residual standard deviation were found using SL anomalies. The results were even larger when short periods (a year or less) were utilized, which is similar to the time period used to compute SSB models shortly after launch. This study emphasizes that the SSB is perhaps the largest current contributor to SSH error and is still a topic of ongoing research.

3.05.2.5 Inverted Barometer Correction

The inverted barometer (IB) correction is not a correction to the raw range measurement per se, but a correction to the SSH measurement to account for SL variations caused by changes in the atmosphere pressure. A 1 mbar increase in the atmospheric pressure causes a decrease of approximately 1 cm in the sea surface. Because there is no net pressure gradient at the sea surface due to this atmospheric loading, the signal must be removed from the SSH data before using it to compute geostrophic currents so as not to corrupt the calculation. The size of the variation is also large in some regions, up to 15 cm or more in the Southern Ocean. Because height variations are large and not due to internal ocean dynamics, they are frequently removed using a model that assumes an instantaneous, static response to surface atmospheric pressure (P_{atm}):

$$\mathrm{DSSH_{IB}} = a(P_{atm} - P_{ref}) \qquad [14]$$

The correction is subtracted from the SSH measurement. Different scaling factors (α) have been used for different altimeters, based on interpretations of various static IB studies (e.g., Dickman, 1988; Wunsch, 1972). For example, the Jason-1 GDRs use a value −9.948 mm mbar^{-1}, while the T/P and Geosat GDR handbooks recommend a value of −10.1.

More problematic is the different recommendations for the reference pressure in the GDR handbooks. Even though Ponte et al. (1991) noted that the reference pressure should be the average pressure over the ocean as a function of time (see also Dorandeu and Le Traon, 1999), the reference pressure used for the Geosat and T/P GDRs was a fixed mean pressure based on the global average (1013.3 mbar), not the time mean over only the ocean (~1010.9 mbar). It also did not include the time variation, which has a seasonal amplitude of about 0.6 mbar and intraseasonal variations of nearly the same magnitude (Dorandeu and Le Traon, 1999). One artifact of this was that the GMSL calculated from T/P data with the IB correction in the GDRs applied had an erroneous seasonal variation (Dorandeu and Le Traon, 1999). It also means that there is an approximately 24 mm bias difference between SSH computed with the different reference pressures.

Jason-1 GDRs use the correct IB response, calculated relative to the instantaneous mean ocean pressure derived from the ECMWF atmospheric model, that is,

$$\mathrm{DSSH_{IB}} = -9.948(P_{atm} - P_{ref}(t)) \qquad [15]$$

3.05.3 Satellite Orbit

3.05.3.1 Orbit and Ground Track

The subsatellite point of any object orbiting the Earth will trace out a path referred to as the ground track when the orbit is projected from an inertial reference frame (i.e., one with a fixed orientation in space) to an Earth-fixed reference frame (i.e., one with a fixed orientation relative to the Earth). To first order, the two frames can be considered to have the same center and z-axis along the pole, but the Earth-fixed frame rotates about the inertial frame at a rate ω_E where

$$\omega_E = \frac{2\pi}{86164s} \qquad [16]$$

Note that the period of rotation is equivalent to the length of a sidereal day (expressed in solar seconds), not a solar day. Because of this, after every revolution of the satellite in its orbit, the subsatellite point will cross the equator westward by an amount equivalent to the angle that the Earth has rotated during that time

$$\mathrm{D_E} = w_E P \qquad [17]$$

For a circular orbit, the orbital period (P) is given by

$$P = 2\pi\left(\frac{a^3}{\mu_E}\right)^{1/2}\left[1 - \frac{3}{2}J_2\left(\frac{R_E}{a}\right)^2\left(4\cos^2 i - 1\right)\right] \qquad [18]$$

and is dependent on orbit's inclination relative to the equatorial plane (i), the mean radius of the orbit (a), the gravitational constant of the Earth (μ_E), the Earth's equatorial radius (R_E), and the second zonal harmonic of the Earth's gravity field (J_2), which perturbs the period slightly from that predicted by a point mass assumption. For typical satellite altimeter orbits, the orbital period will be about 90 min, which means that the subsatellite point will cross the equator on the next revolution approximately 22.5° west of the previous point.

We have already included the effect of the Earth's equatorial bulge (J_2) in the calculation of the orbital period. However, the same perturbation causes the angular momentum vector of the orbit to precess, like that of a spinning top. The effect of this is

to cause the node of the orbit, or the point here it crosses the inertial equatorial plane, to move either east or west. The node rate ($\dot{\Omega}$) for a circular orbit is determined from

$$\dot{\Omega} = -\frac{3}{2} J_2 \left(\frac{\mu_E}{a^3}\right)^{1/2} \left[\frac{R_E}{a}\right]^2 \cos i \qquad [19]$$

For altimeter satellite orbits, the value for $\dot{\Omega}$ is $\pm 3^\circ$ per solar day. The effect of the node rate acts opposite of the axis rotation for inclinations less than 90° and in the same direction for inclinations greater than 90° so the total ground track shift (Δ) is

$$\Delta = (\omega_E - \dot{\Omega})P \qquad [20]$$

and positive values mean a shift of the ground track westward (Figure 7).

Many different types of ground tracks can be found by slightly adjusting either the radius of the orbit or the inclination to cause a change in Δ. One special type of ground track that is commonly used for satellite altimeter missions is an exact-repeat round track, or orbit. An exact-repeat orbit has a ground track that repeats itself (to within a very small difference) after a certain period, P_R. For altimetric studies, this is desirable for several reasons. For one, the satellite will repeat measurements at a consistent interval, which is important for examining the time-varying SSH. Secondly, by repeating measurements over the same area, one can obtain a precise average of the SSH that is not affected by large time-variable SSH signals such as tides and ocean eddies, which can have variations in SSH of more than a meter. Finally, there is also a need to make measurements in an area as small as possible. Otherwise, the sampling of stationary geoid gradients (of order 5–10 cm km^{-1}) in time could be interpreted as real SSH variation even if there has been no change.

One major drawback to an exact-repeat orbit is that portions of the ocean are still never measured due to gaps in the ground track coverage (Figure 1). Given enough time, a satellite altimeter in a nonrepeating ground track will eventually observe the SSH at a scale approaching the along-track sampling rate, but perhaps only once for any particular area. Although there is the potential to bias the measurement due to large time-variable features like an eddy, the error introduced may not be significant compared with the size of the desired signal. This was one of the reasons that the original Geosat mission was in a nonrepeating ground track, as the U S Navy was more interested in obtaining any measurement of the SSH in areas that had never been observed in order to detect seamounts, trenches, and undersea ridges than the most accurate mean measurement.

A Sun-synchronous orbit is a special type of exact-repeat orbit (Mazzega, 1989). This type of orbit has a node rate $\dot{\Omega} = 2\pi/365.2422$ days, which means that the angle between the orbital plane and the vector between the Earth–Sun line will maintain a constant angle. This is beneficial in that the satellite will fly over a region at roughly the same local time for every ascending or descending pass and can be useful for maintaining a fixed orientation of the solar panels with the Sun for charging. Although this type of orbit is usually used for satellites that require a consistent viewing angle relative to the Earth–Sun line, there have been three altimeter satellites in Sun-synchronous orbits: ERS-1, ERS-2, and Envisat. These satellites were hosts to instruments other than the radar altimeter, and these required the Sun-synchronous orbit. There are certain aspects of Sun-synchronous orbits that will be discussed further in Section 3.05.3.2 that make them problematic for some applications of altimetry.

For an exact-repeat orbit, there are an integer number of orbital revolutions (N) that are equivalent to D nodal days. A nodal day is the time required for the Earth to make one rotation relative to the satellite orbit's line of nodes. Thus, to find the exact-repeat orbit, one needs to solve the equation

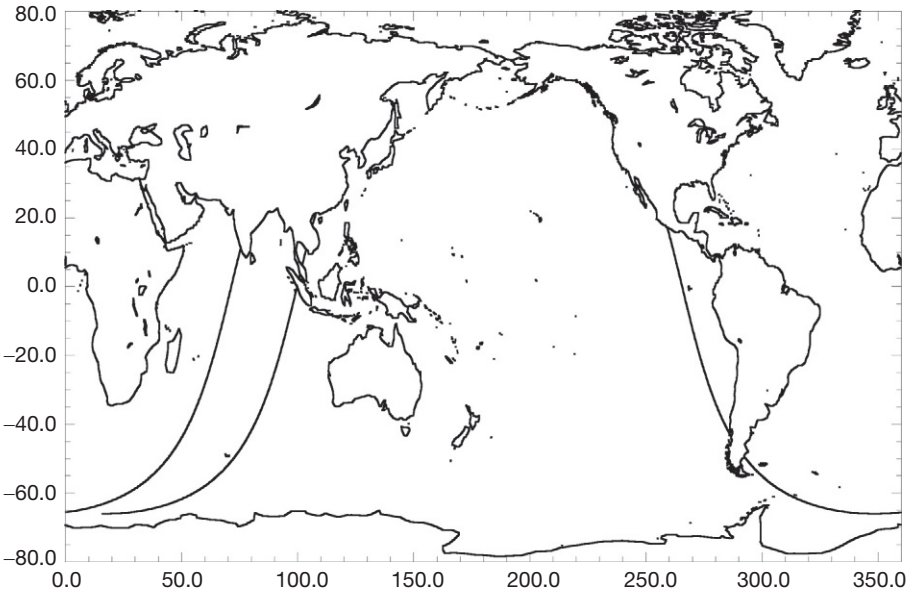


Figure 7 Plot for T/P ground track for first three passes, showing westward shift of ground track in time.

$$P_R = NP = \frac{2\pi D}{(\omega_E - \dot{\Omega})} \quad [21]$$

to find the orbit radius (assuming a circular orbit) and inclination that gives one a 'D-day' repeat orbit. Equation [21] is nonlinear so there is no analytic solution. Typically, one starts with an approximate orbital radius and inclination, calculates an initial P and $\dot{\Omega}$, then assumes a value for D, and calculates N (either rounding or truncating to make it an integer). Now that an approximation of N is known, one can recompute the orbital period, adjust either the radius or inclination (usually radius) to match the period, and then iterate until the solution converges. For Sun-synchronous orbits, P_R is exactly an integer number of days equal to D. For other exact-repeat orbits, P_R is a value slightly shorter or larger than D. For example, T/P (with an inclination of $\sim 66^\circ$) is called a '10-day repeat orbit' (the value of D), but the actual repeat period is 9.915625 solar days. Geosat has a '17-day repeat period' but the actual period is 17.0505 days.

The spacing (Δx) between adjacent ascending and descending passes of the ground track at the equator is determined from the number of revolutions in the repeat period by

$$\Delta x = R_E \left(\frac{2\pi}{N}\right) \quad [22]$$

Thus, longer repeat periods (and hence higher N) will have shorter distances between ground tracks. T/P has 127 revolutions in the 10-day repeat period, with a ground track spacing of ~ 315 km, while Envisat has 501 revolutions in a 35-day repeat and a spacing of ~ 80 km. Note that ground tracks that are adjacent in space on the Earth's surface are not adjacent in time. For instance, although the spacing of ascending tracks for T/P is $\sim 2.8^\circ$, the track immediately after the previous one in time is 28.3° westward at the equator (e.g., Figure 7).

The ground track will not remain stable due to nonconservative forces such as drag that removes energy from the orbit. The effect of drag is to reduce the semimajor axis of the orbit. This means that the orbital period will also decrease (eqn [18]) and that the ground track shifts (Δ) will also decrease (eqn [17]), so that the ground track will drift eastward relative to its nominal location (Figure 8). To stop the drift and reverse it, the semimajor axis must be raised via a pair of orbital maintenance maneuvers in order to increase the period and cause a drift back to the west. In order to not only maintain an exact-repeat ground track but also not expend too much fuel, the drift is allowed within a certain band, generally ± 1 km from nominal for satellite altimeter missions. With some knowledge of the atmospheric density and drag properties of the spacecraft, one can effectively steer the ground track to stay within the ± 1 km range (Figure 8).

Note that the inclination of the orbit also determines approximately the maximum latitude of the ground track:

$$\begin{aligned} f_{max} &\approx i \text{ for } \quad i < 90^\circ \\ f_{max} &\approx 180^\circ - i \text{ for } \quad i > 90^\circ \end{aligned} \quad [23]$$

3.05.3.2 Aliasing

One consequence of the relatively long intervals between successive samples at the same location due to the satellite orbit is aliasing of signals with periods smaller than the exact-repeat period into signals with periods significantly longer than repeat

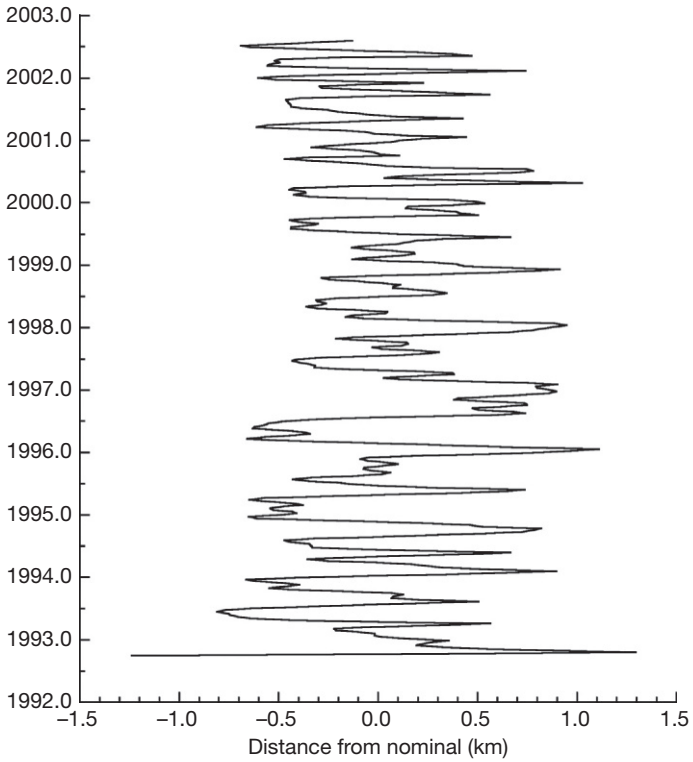


Figure 8 Plot of T/P ground track variation around nominal ground track at equator for duration of primary mission until it was moved to its current ground track. Positive excursions indicate the position is east of nominal, while negative values indicate the position is west of nominal.

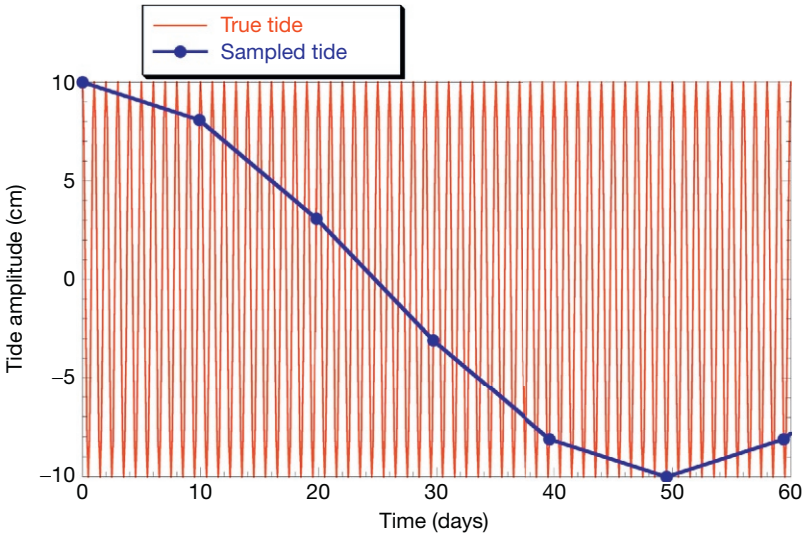


Figure 9 Aliasing of a 1 cycle-per-day signal (red) by sampling every 9.9 days (blue).

period. This is demonstrated in Figure 9, where a variation with a frequency of exactly 1 cycle per day (like an ocean tide) is sampled every 9.9 days. Here, the altimeter sees the tidal variation as a low-frequency fluctuation (period = 99-days), although with the same amplitude. This means that if the high-frequency variation is not modeled and removed from the data properly, errors will leak into the SSH measurement as long-period signals that could be interpreted as real SL variations.

Ocean tides are one such high-frequency signal. The dominant tidal frequencies occur in bands near one (diurnal) and two (semidiurnal) cycles per day. Most of the periods are slightly different from 12 and 24 h (Table 1), except for the S2 tide (caused by solar perturbations), which has a period of exactly 12 h. This has an important consequence for a Sun-synchronous orbit. Because the exact-repeat period for a Sun-synchronous orbit is exactly an integer number of N solar days, it will always sample the S2 tide at exactly the same phase every day. Thus, the tide (or tide error) will alias into a bias, and not a long-period variation. The alias period may also be so long that it is essentially a bias, depending on the length of the mission.

The theory of tidal aliasing for altimetry is described by Parke et al. (1987). Given the exact-repeat period, P_R, one computes the tidal phase change over one repeat period by

$$\Delta\vartheta_{\text{tide}} = \frac{2\pi P_R}{P_{\text{tide}}} \quad (\text{angle between} -\pi \text{ to } \pi) \qquad [24]$$

where P_{tide} is the period of the particular tide (Table 1). Both periods (P_R and P_{tide}) must be in the same units. The resulting principal alias period is

$$\tau_{\text{tide}} = \text{abs}\left(\frac{2\pi P_R}{\Delta\vartheta_{\text{tide}}}\right) \qquad [25]$$

Table 1 gives the alias periods at the major tidal constituents for the various altimeter exact-repeat orbits. Note that in addition to the bias at S2 for the Sun-synchronous orbits, the periods of the K2, K1, and P1 tides are close enough to 12 and 24 h that they alias into very long periods near the semiannual period (K2) and annual periods (K1 and P1). The P1 tide for Geosat orbits aliases into a 12-year fluctuation, which is nearly a bias over short periods. The alias periods for T/P are all less than 180 days, which was by design (Parke et al., 1987) in order that the data could be used to improve tide models (see Section 3.05.5.2).

Aliasing would not be important if the tidal variations were known accurately. However, this was not the case, even as recent as a decade ago. Tide models developed with Geosat data (e.g., Cartwright and Ray, 1990) had errors greater than 5 cm RMS in the open ocean. Although this was large, it was still better than errors in older tide models based only on tide gauge data or hydrodynamic models. However, with only a year of T/P altimeter data, a reduction in error to the level of 2–3 cm was possible, due mainly to the short alias periods (e.g., Ma et al., 1994; Schrama and Ray, 1994). Although tide errors are now on the order of a few centimeters RMS in the open ocean (Fok et al., 2010), any remaining error in the model for a certain constituent will still alias into a small, undesired long-period signal.

In addition to temporal aliasing, tides will be spatially aliased as well (Parke et al., 1998; Schlax and Chelton, 1994). Because tidal signals propagate across the ocean as long waves and there is a several-day interval between adjacent ground tracks, tidal errors will also propagate as waves. This can cause spurious signals that may be interpreted as real Kelvin or Rossby waves at certain latitudes. This was observed in Geosat data with the Cartwright and Ray (1990) tide model (Jacobs et al., 1992) and also in T/P before tide models were improved (Schlax and Chelton, 1994). With current tide models, the data from the T/P ground track now show no significant spatial aliasing and have been used to show for the first time Rossby waves at most latitudes (Chelton and Schlax, 1996).

There are SSH variations other than tides with periods of less than 10 days, which will also alias into long-period errors. More problematic is that these variations are not regular like tides; both the amplitudes and phases will shift in time. Fukamori et al. (1998) used a general ocean circulation model forced by realistic winds and heat fluxes to demonstrate that half the energy at periods of less than a year in some regions was due

Table 1 Periods of largest tidal constituents along with altimeter alias periods (in days)

Tide symbol	*Period (h)*	*Equilibrium amplitude (cm)*	*Source*	*T/P 10-day alias*	*Geosat 17-day alias*	*Envisat 35-day alias*
M_2	12.420601	16.83	L	62.1	317.1	94.5
K_1	23.934470	9.83	C	173.2	175.5	365.2
S_2	12.000000	7.83	S	58.7	168.8	∞
O_1	25.819342	6.99	L	45.7	112.9	75.1
N_2	12.658348	3.26	L	49.5	52.1	97.4
P_1	24.065890	3.25	S	88.9	4466.7	365.2
K_2	11.967235	2.13	C	86.6	87.7	182.6
Q_1	26.868357	1.35	L	69.4	74.1	132.8

Source: L = lunar, S = solar, C = combined.

to barotropic variations with periods of less than 20 days. Such variations are not properly sampled by altimeters (unless they have very short repeat periods) and so will alias into long-period errors. The barotropic variations are predominantly at higher latitudes and have amplitudes as large as 5 cm.

Fukamori et al. (1998) suggested that barotropic variations from a model could be used to dealias the altimetry data, and this has been tested (e.g., Stammer et al., 2000; Tierney et al., 2000). The variance of the T/P SSH was reduced by about 10–20% in regions of high barotropic variations, suggesting that the model was removing the signal from the altimetry. Because of these studies, a correction to reduce high-frequency aliasing based on an ocean model is available on Jason-2/OSTM GDRs as a standard, recommended correction.

3.05.3.3 Precise Orbit Determination

Determining the position of a satellite accurately at any time depends on having an accurate dynamic model in order to propagate the orbit forward in time. Generally, the orbit will be in error for several reasons. The initial starting position may be wrong. Parameters in the dynamic models – such as the Earth's gravitational acceleration, atmospheric drag, and Earth and solar radiation pressure – may be incorrect. If there are independent measures of the satellite position or velocity, then conceptually, one can use the difference between the observed position and the computed position to estimate corrections to the dynamic models and initial position in order to reduce the differences. In practice, there is no way to measure the satellite position or velocity directly, but instead, one measures either the range from another site or the change in rate of the range. For a complete analysis of POD techniques, please see Tapley et al. (2004a) or, as they apply to altimetry, Chelton et al. (2001). Here, we wish to summarize the major improvements in POD that have led to more accurate measures of SSH from altimetry and how orbit errors appear in the SSH. Since the errors in SSH are directly proportional to errors in the nadir altitude of the satellite (eqn [1]), one is most concerned with errors in what is referred to as the radial component of the orbit. Other components of the orbit (defined in a satellite-fixed frame) are the transverse component (in the direction of satellite motion) and the normal component (normal to the radial and transverse in a right-hand sense). Errors in the transverse and normal component affect the SSH much less than errors in the radial component, although they will affect the accuracy of positioning the subsatellite point, but not significantly compared to the width of the radar footprint. In the literature (as here), when altimeter orbit accuracy is discussed, radial orbit accuracy is implied unless otherwise stated.

Initial orbits for both Seasat and Geosat had errors of several meters (Born et al., 1988; Marsh and Williamson, 1980; Schutz and Tapley, 1980). This was significantly larger than the total SSH accuracy goal for the T/P mission (13 cm), a value that was still higher than the 5 cm or better value requested by oceanographic community (Stewart et al., 1986). However, the dominant signature of orbit error is that of a very long-wavelength once-per-orbital revolution sinusoid, due mainly to dynamic errors in the orbit introduced by errors in the gravity field model (e.g., Tapley et al., 2004a). There are also two components of the orbit error. One varies in time so that the error is different from one pass to another. The second component is phase locked to a specific geographic location and will tend to be the same over time and is known as geographically correlated orbit errors (Tapley and Rosborough, 1985).

Geographically correlated orbit error is aliased directly into any calculation of the MSS (**Section 3.05.5.1**). However, one can effectively remove this correlated orbit error from the altimeter data by estimating and removing the mean SSH from the data and examining only the time-variable component. The time-variable component of orbit error will remain, however. Several methods have been devised to reduce the time-variable portion of the orbit error in the data. Because it is a long-wavelength signal, a bias and trend was often fit over portions of each pass of SSH anomalies (SSHA) and the residuals were examined. Although this meant one could only study wavelengths shorter than the pass length utilized, it did allow for a reduction of the orbit error from order 1 m to order 10 cm (Tapley et al., 1982). Tai (1988) used a sinusoid with a frequency of 1 cycle per revolution (cpr) and estimated the amplitude and phase over each satellite orbital revolution (2 passes). Other studies have shown similar results using complex demodulation (Francis and Bergé, 1993), using a Fourier series (Douglas et al., 1984), estimating a sinusoid over each pass (half revolution) (Chelton and Schlax, 1993), or using an adaptive Kalman filter (Urban, 2000).

The one problem with all these techniques is that they will absorb some of the real long-wavelength ocean variability in addition to the orbit error (Van Gysen et al., 1992). This is more likely when the SL variation is of the same order of magnitude as the orbit error. One of the more important long-wavelength

ocean signals that may be removed is the annual variability in the SL due to heating/cooling of the upper layers of the ocean at midlatitudes and wind forcing in the tropics. In general, when the sea surface in the Northern Hemisphere is higher than the time average, the sea surface in the Southern Hemisphere will be lower than average (**Figure 10**). The difference from peak to trough over a pass can be as large as 15–20 cm, and since the satellite circles the Earth once every revolution, this real ocean signal will tend to alias into a 1 cpr signal in the spectra of the satellite SSHAs (Wagner and Tai, 1994). Furthermore, if a sinusoid is fit over longer arcs (i.e., multiple revolutions), it will not remove the total orbit error and could absorb signals from other sources, since the amplitude and phase of the residual orbit error change in time. Thus, for many oceanographic applications, orbit errors in the range of 10–20 cm are more problematic than errors larger than 50 cm.

As noted previously, the largest contributor to orbit error are errors in the Earth gravity model, followed closely by errors in drag and solar radiation pressure. The T/P orbit was designed to have a greater semimajor axis than either Geosat or Seasat in order to reduce both the effects of drag and errors in the Earth gravity field model at the time, as gravity errors (like the signal) are attenuated with distance. However, predicted orbit errors from the best gravity field models at the time, like the Goddard Earth Model (GEM) GEM-10B (Lerch et al., 1981), were still estimated to be of the order of 50 cm (Nerem et al., 1994a). Thus, significant work was done in order to improve the gravity field models, using in part altimeter data from the Geosat mission. These efforts led to the development of the Joint Gravity Model-1 (JGM-1) by members of the T/P SWT, which was the prelaunch model for T/P (Nerem et al., 1994a). The orbit accuracy using JGM-1 along with other model improvements (Tapley et al., 1994a) significantly exceeded the goal of 13 cm, reaching an estimated radial accuracy of about 5 cm (Nerem et al., 1994a). However, further improvements to the gravity field model (JGM-2) were made, mainly from the inclusion of T/P tracking data (Nerem et al., 1994a). The RMS of the orbit error using JGM-2 was estimated to be about 2 cm (Tapley et al., 1994a), with total orbit errors (including errors in models other than gravity) of about 3 cm RMS. A further iteration of the JGM model resulted in JGM-3 (Tapley et al., 1996). Total orbit errors for T/P using JGM-3 are estimated to be 2 cm RMS or less (Tapley et al., 1996). Although improvements in the gravity models had the largest effect in improving T/P orbits, improvements in other models also contributed to the overall orbit accuracy. These included a model for the satellite's complex shape along with its thermal and reflective properties, better solid and ocean tide models, and improved empirical parameterization (e.g., Tapley et al., 1994a).

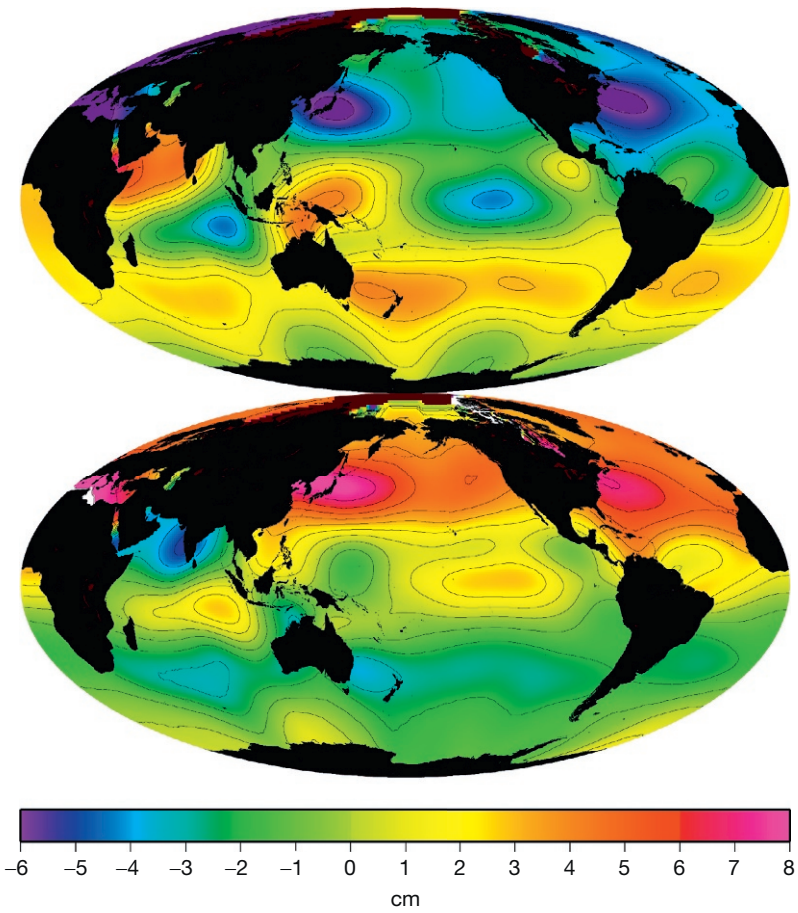


Figure 10 SSH anomalies for average March (top) and average October (bottom) based on 12 years of T/P and Jason-1 maps that have been smoothed with a 1000 km radius Gaussian.

It was still difficult to match this accuracy with other satellites such as ERS-1, ERS-2, and GFO until recently because they are at a lower altitude and hence have higher drag and are sensitive to gravity signals not completely modeled by JGM-3. They have also generally had less tracking data than T/P; the PRARE system failed on ERS-1, and only limited SLR tracking was available. Although both PRARE and SLR tracking were available for ERS-2, PRARE provides significantly less tracking data than the DORIS system on T/P. The GPS system failed on GFO and it too had to rely only on SLR tracking. However, by including tracking data from ERS-1 and ERS-2 and computing new gravity models tuned to each particular satellite (Scharroo and Visser, 1998; Tapley et al., 1997), orbit accuracy for these low-altitude satellites was improved, although still not to the level of T/P. For example, the fit of SLR residuals to the precise orbit, which is one indicator of orbit improvement, was reduced from order 7–8 cm RMS to 3–4 cm RMS using these new gravity models (Scharroo and Visser, 1998; Tapley et al., 1997). Typical values for T/P are 1–2 cm RMS.

A significant improvement in orbit determination of low-altitude satellites is now possible because of an improved gravity model determined from the Gravity Recovery and Climate Experiment (GRACE) (Tapley et al., 2004b, 2005). For example, Lemoine et al. (2003) computed orbits for the low-altitude Envisat satellite using both SLR and DORIS tracking and a variety of gravity models. They found that the SLR residuals for orbits computed with the EGM96 gravity field model had an RMS of 4.8 cm, but when an early GRACE gravity model was substituted, the residual RMS was reduced to 2 cm, which is comparable to the best orbits for T/P.

Accuracy for Jason-1, which is in the same orbit as T/P, is generally even better than that of T/P, having estimated errors of 1 cm RMS or less (Choi et al., 2004; Luthcke et al., 2003). This is due primarily to the use of GPS tracking data in addition to SLR and DORIS. The use of new GRACE gravity models does not significantly alter the time-variable orbit error for T/P or Jason-1, although it does reduce the geographically correlated portion (Beckley et al., 2004; Choi et al., 2004). Jason-2/OSTM carries a BlackJack GPS receiver similar to Jason-1. Standard orbits on the Jason-2/OSTM GDRs are comparable to those for Jason-1. However, POD methods have progressed, and using special algorithms to reduce phase biases, it has been shown that orbit accuracy can be reduced further to a value approaching 6 mm RMS (Bertiger et al., 2010).

The difference between a 2 cm orbit and a 1 cm orbit may not seem large, but the effects in terms of SSH error are significant for long wavelengths. Recall that the error is predominantly at 1 cycle per revolution. As Chelton et al. (2001) pointed out, a 2 cm orbit is equivalent to an erroneous rise (or fall) in SSH of about 4 cm over a distance of 10 000 km, the length scale of a typical ocean basin. This can still be a significant fraction of the basin-scale SSH variation (especially that related to the seasonal heat cycle).

3.05.4 Calibration and Validation

It is vital to calibrate and validate the SSH measurement after all the corrections discussed in Section 3.05.2 are applied. The term calibration/validation (cal/val) is often used in the literature, but one should understand that each term is very distinct. The altimeter SSHA (variation about a mean value) is often validated against independent measures of the SSH variation, typically from tide gauges but more recently from other satellite altimeters. By comparing SSHAs, one is not concerned with either biases in the altimeter or the independent data. Typically, validation consists of determining if the frequency and amplitudes of the observed SSHAs are similar, usually by computing the correlation and standard deviation (or RMS) between the two time series after means have been removed. High correlation and low RMS are interpreted as meaning the altimeter is accurately measuring SSH. Often, the altimeter accuracy in measuring SSH is given from the RMS of the comparison (or mean RMS averaged over several sites) assuming that the tide gauge has no errors, or as RMS/$\sqrt{2}$ assuming that the error in the tide gauge is comparable to that of the altimeter and there are no correlated errors.

Calibration, on the other hand, refers to the comparison of absolute SSH, not anomalies. In this case, one assumes that the SSH at an independent site has been measured in the same reference frame as the altimeter uses (i.e., relative to the same local ellipsoid) and is free of biases. In this case, one can measure the mean offset between the altimeter SSH measurement and the independent 'truth' measurement and interpret it as a bias in the altimeter SSH (or range). The change in the bias over time can also be tracked using this method, although it can also be tracked using comparisons to tide gauges that are not in a known reference frame (see Section 3.05.4.2).

In the following sections, we will discuss the methods of altimetry cal/val and comment on the strengths and weakness of each method. We will also discuss the results of the first direct calibration between two altimeters in the same orbit and at about the same time, when T/P and Jason-1 were observing the same SSH separated in time by only 70 s as part of the Jason-1 cal/val phase. In particular, we will comment on several insights that this unique study provided on cal/val methods.

3.05.4.1 Calibration at Single Site, Instrumented Platforms

The oldest cal/val method used in altimetry is based on comparing the measurement directly to an independent measure of SSH made at a site that the satellite directly overflies. This was done with both GEOS-3 and Seasat (Kolenkiewicz and Martin, 1982; Martin and Kolenkiewicz, 1981) and was successful even in the presence of meter-level orbit errors because the overflight site at Bermuda also had a NASA laser ranging station. Because of this, the orbital altitude could be determined very precisely at the calibration point. The altimeter bias is determined by comparing the SSH with that measured by one or more tide gauges at the calibration site. However, tide gauges normally measure SSH variation relative to some arbitrary level that is not defined in the same reference frame as the altimeter. Thus, one first needs to find the tide gauge reference level relative to the measurement frame used by the satellite (i.e., relative to the same reference ellipsoid). If that is done accurately, then the 'true' absolute SSH at the tide gauge can be obtained. Then the altimeter SSH measurement can be compared to the 'truth,' and the difference is interpreted as the bias. Generally, multiple observations are made over a period of months or years to average out random errors.

One problem with this method was that neither Seasat nor Geosat flew directly over the tide gauge, which meant that the altimeter SSH measurement had to be interpolated to the tide gauge. In addition, the survey of the tide gauge relied on traditional leveling techniques from a more distant benchmark, which may have introduced systematic errors at the several centimeters level. However, the method was sufficient for bias determination in the presence of meter-level SSH accuracy.

Geosat had no SLR reflector and so no direct calibration of the measurement was undertaken, since the primary mission was more interested in SSH gradients than absolute SSH. This is one significant problem with trying to link Geosat and T/P data in order to determine longer time series of GMSL change (**Section 3.05.3**), although it has been attempted using uninstrumented tide gauge sites (Guman, 1997; Urban, 2000). However, the uncertainty in such calibrations is of the order of several centimeters.

ERS-1 SSH was calibrated over a nearly a 2-month period in 1991 using a method similar to that used for GEOS-3 and Seasat (Scharroo, 2002). There were several significant improvements in the campaign. First, the site was located on a small scientific platform in the Adriatic Sea off the coast of Venice. This meant that there was no loss of signal due to land as was the case with Bermuda. In addition, the ground track for ERS-1 was selected to fly directly over the platform. The average distance between the nearest 1s altimeter point and the tide gauge was only 600 m. Finally, and perhaps more importantly, GPS was available to tie the tide gauge to the ERS-1 reference frame. This had the potential to significantly reduce systematic bias errors in the 'truth' measurement. The initial estimate indicated an altimeter range bias of +4.5 m (Scharroo, 2002). However, after 2 years of analysis and many corrections to the data (including four changes to the reference level of the tide gauge), the final range bias estimate was −41.5 cm, based on ten overflights with values ranging from −46.9 to −21.6 cm (Scharroo, 2002).

The T/P mission used two instrumented calibration sites, the main one off the coast of California on the Harvest oil platform (Christensen et al., 1994) and the other in the Mediterranean Sea on Lampedusa island (Ménard et al., 1994). The tide gauges were also leveled to the same reference frame as T/P using GPS. Like the ERS-1 calibration site, the Harvest platform was specifically chosen because it was very nearly under one pass of the T/P ground track so that interpolation was minimized. Unlike the ERS-1 site, the Harvest platform was designed to be a continuing calibration site and has continued to do so to this day (Haines et al., 2003, 2010). Because of this, the initial calibration at Harvest was based on nearly a year's worth of overflights. The scatter of the bias estimates was also smaller than those for ERS-1, with a standard deviation of about 2 cm. The Lampedusa site, on the other hand, was designed for a limited campaign and in fact only estimated a TOPEX bias based on two overflights (Ménard et al., 1994). The initial bias estimate for TOPEX was about 15 cm (Christensen et al., 1994; Ménard et al., 1994). It was eventually discovered that an oscillator correction had been improperly applied to the TOPEX data (e.g., Nerem et al., 1997). When this was corrected, the apparent TOPEX bias disappeared.

With the advent of relatively inexpensive GPS receivers, a new type of calibration instrument has been developed: the GPS buoy. The idea behind a GPS buoy is relatively simple. Data from a GPS receiver mounted on a floating buoy can be used to determine the local SSH in the same frame as the altimeter. If the buoy is deployed along a ground track at the same time as the satellite altimeter is flying overhead, one can compare the two SSH measurements and determine the bias. Born et al. (1994) demonstrated the concept using a buoy at the Harvest platform. They found that the bias determined from the GPS buoy for a single over flight agreed with the value determined at the Harvest platform to within 1 cm.

Based on the success of both the Harvest tide gauge and GPS buoy campaigns, several other calibrations sites were devised for the Jason-1 mission, many of them using a combination of GPS-referenced tide gauges and GPS buoys. In addition to the Harvest platform, they ranged from a network of tide gauge sites in the United Kingdom (Woodworth et al., 2004) to sites at several islands in the Mediterranean Sea (Bonnefond et al., 2003; Jan et al., 2004; Martinez-Benjamin et al., 2004; Pavlis and Mertikas, 2004) and one site in Tasmania (Watson et al., 2003).

Although the goal of these calibrations was to quantify the absolute altimeter range bias of Jason-1, realistically, only the local bias in SSH is estimated. The bias estimate will be corrupted by biases due to orbit errors, SSB error, or the water vapor correction, all of which can have large geographically correlated biases. The large number of absolute calibrations sites for Jason-1 allows one to better quantify some sense of the true error in this type of absolute calibration (**Table 2**). Although formal errors on all the estimates (based on scatter and number of observations) are of the order a few millimeters, the absolute calibration estimates differ by up to several centimeters. This result suggests that realistically, it is difficult to determine the absolute bias to better than the 1–2 cm level using these single-site calibration techniques.

3.05.4.2 Cal/Val from Comparison to Tide Gauges

One can compute SSHAs from both altimeter and tide gauge measurements (even if they have not been consistently referenced) and compare the results. This is typically done by comparing the closest SSHA along altimeter passes that bound a tide gauge to the daily average tide gauge measurement (e.g., Mitchum, 1994) or by averaging altimeter data from multiple passes around the site over some period (typically one repeat cycle or a month) and comparing it to the tide gauge data averaged over the same time period (e.g., Chambers et al., 1998a; Cheney et al., 1994). The former can be thought of as a measure of the error in the 1s SSH measurement, while the latter is an estimate of some smoothed SSHA.

Wyrtki and Mitchum (1990) found RMS differences of 3–10 cm between tide gauge data and Geosat data averaged into 2° latitude by 8° longitude grids. Cheney and Miller (1990) and Shum et al. (1994) obtained similar results using different orbits and processing. Mitchum (1994) compared TOPEX and tide gauge data early on in the mission and found that the two agreed with each other on the order of 5 cm RMS, which is more representative of the error in a single 1s value because of the methodology employed. Other authors who used temporal and spatial averaging found RMS errors in the range of 2–4 cm (Cheney et al., 1994; Nerem et al., 1994b; Tapley et al., 1994b). Chambers et al. (1998a) analyzed 3 years of TOPEX data around 40 tide gauges in the Pacific and Indian Oceans and

Table 2 Bias estimates for Jason-1 using original GDRs

Site/technique	*Absolute bias*	*Relative bias with TOPEX_B*	*Reference*
Harvest	13.8	11.9[a]	Haines et al. (2003)
Corsica	12.0	10.1[a]	Bonnefond et al. (2003)
Bass Strait	13.1	11.2[a]	Watson et al. (2003)
Corsica	11.8	9.9[a]	Jan et al. (2004)
Gavdos, Crete	14.5	12.6[a]	Pavlis and Mertikas (2004)
United Kingdom	12.9	11.0[a]	Woodworth et al. (2004)
Ibiza	12.0	10.1[a]	Martinez-Benjamin et al. (2004)
Mean (std. dev.)	12.9 (1.0)	11.0 (1.0)	
Global Residuals	–	15.3	Beckley et al. (2004)
Global residuals	–	14.4	Chambers et al. (2003a)
Global residuals	–	14.1	Dorandeu et al. (2004)
Global mean sea level	–	15.4	Leuliette et al. (2004)
Global residuals	–	14.6	Shum et al. (2003)
Mean (std. dev.)		14.8 (0.6)	

Bias in cm of SSH, computed so $SSH_{true} = SSH_{meas} - bias$.

[a]Relative bias computed relative to TOPEX_B absolute bias of 1.9 cm measured at Harvest (Haines et al., 2003).

found a mean RMS of 3.5 cm and an average correlation of 0.87. During the Jason-1 cal/val phase, Chambers et al. (2003a) used the unique fact that T/P, Jason-1, and the tide gauges were observing the SL variations at roughly the same time in order to quantify the mean error in each instrument, based on an assumption that errors between the three data sets were uncorrelated. They estimated errors in the TOPEX, Jason-1, and tide gauge SSHAs at 3.1, 3.7, and 1.6 cm, respectively. The larger error in Jason-1 is largely attributable to the fact that the onboard tracker for Jason-1 does less smoothing to the waveforms than the TOPEX hardware, so the range measurement is slightly noisier than that of TOPEX.

Although one cannot determine the 'absolute' altimeter bias with the general tide gauge network, one can use the data to determine changes in the bias. The strength of this is that averaging altimeter–tide gauge residuals at many sites will effectively reduce the formal error in the determination, compared to the error at a single site such as the Harvest platform. This means that one has the potential to detect even small changes in the bias, of the order of several millimeters. Although one can simply average altimeter–tide gauge residuals to reduce the noise (e.g., Chambers et al., 1998a), the variance can be reduced even more if one accounts for both land motion at the tide gauges and phase differences between the altimeter and tide gauge measurements due to propagating ocean waves (Mitchum, 1998, 2000). For example, Haines et al. (2003) found that the TOPEX residuals at the Harvest calibration site were about 3–4 cm RMS, while Mitchum (1998) found that the RMS of the residuals averaged over 55 tide gauges was less than 1 cm.

The first significant use of this method was to detect a large, quadratic trend in the first release of the TOPEX SSH data, which was causing TOPEX SL to differ from that measured by tide gauges by nearly 7 mm $year^{-1}$ (Mitchum, 1998). The source of the error was eventually traced to an improperly applied oscillator correction, which caused not only the drift but also the order 15 cm bias first detected in the TOPEX data (Christensen et al., 1994; Ménard et al., 1994). However, even after the oscillator error was corrected, the tide gauge calibration continued to suggest a drift in the TOPEX measurement, of the order of 1–2 mm $year^{-1}$, with the sense being that the SL rate measured by TOPEX was not as large as that measured by the tide gauges (Chambers et al., 1998a; Mitchum, 1998). The error was eventually linked to drifts in the radiometer (Keihm et al., 2000; **Section 3.05.2.3**). Note, however, that the official T/P GDRs have still not been corrected for this error and users have to either apply an ad hoc correction based on the comparison of the average TMR path delay with that from other radiometers (**Section 3.05.2.3**) or use a separate correction product, which has been distributed with the GDRs.

After correcting for the TMR drift, another apparent bias change in the TOPEX data was noted shortly after the redundant TOPEX_B altimeter was turned on because of degradation in the original TOPEX_A altimeter. The change was small, of the order of 5 mm, and it took some time to acquire enough data to ensure that the change was significant. Shortly afterward, it was shown that the bias was caused by an error in the SSB model used for the TOPEX_B data (Chambers et al., 2003b; **Section 3.05.2.4**). After applying a new TOPEX_B SSB model, there are no statistically significant drifts or changes in bias in the TOPEX data (Chambers et al., 2003b; Leuliette et al., 2004).

A similar tide gauge calibration continues to be performed on Jason-1 and Jason-2/OSTM. The analysis of the first release Jason-1 GDRs found a significant drift between Jason-1 and the tide gauges of more than 5 mm $year^{-1}$ (Leuliette et al., 2004). However, when a different set of precise orbits were utilized, the drift was reduced to 2.5 mm $year^{-1}$ (Leuliette et al., 2004), and the remainder of the error has been attributed to the bias jumps in the JMR (**Section 3.05.2.3**). When new precise orbits and corrections for the JMR jumps are utilized, the apparent drift in the tide gauge calibration is not statistically different than 0 (Beckley et al., 2010).

Thus, comparing altimeter SSH measurements with tide gauge data has proved to be a very accurate method to detect even small changes in the bias of the altimeter. In all the cases discussed, the bias change was first observed in the tide gauge calibration and then later linked to a specific error. Although calibration at the level of less than 1 mm $year^{-1}$ is not

necessary for many applications, it is vital for determining changes in GMSL, where the signal is only about 2–3 mm year^{-1} (Section 3.05.5.3).

3.05.4.3 The *T/P*: Jason-1 Calibration Phase

ERS-1 and ERS-2 flew along the same ground track for 1 year as part of a calibration phase, but the time separation between the two measurements was 1 day, which meant that the satellites were not observing exactly the same SSH. Although the data were used to calibrate and validate both ERS-1 and ERS-2 measurements (Scharroo, 2002), the results were affected by real changes in the true ocean state during the 24 h time difference.

The Jason-1 calibration phase was planned so that the satellite flew over each point along the ground track approximately 70 s before T/P. Since the ocean and atmospheric signals change much more slowly than 70 s, the TOPEX and Jason-1 science instruments effectively observed identical SSHs in this configuration. This meant that any significant differences (either biases or large variances) between the Jason-1 and TOPEX measurements would be attributable to an error in either one or both of the instruments.

One strength of this cal/val mission was the sheer number of points available for computing statistics, as there are typically more than 400 000 observations considered good for each 10-day repeat cycle. The cal/val phase lasted for 210 days (21 complete repeat cycles), at which time T/P was maneuvered into a new orbit with a ground track halfway between the original ground track. When the cal/val phase was planned, it was assumed that TOPEX was properly calibrated and that any large bias or variance would be attributable to an error in the Jason-1 system. Contrary to expectations, however, early analysis pointed to small, yet evident, problems in the TOPEX data and not the Jason-1 observations. One of these problems was related to the SSB model used for the TOPEX measurements (Chambers et al., 2003b) and has been discussed in **Section 3.05.2.4**. Another problem was related to the wet path delay measured by the TMR, which was found to change during yaw transitions (see **Section 3.05.2.3**). Both of these problems were relatively small and had gone undetected for several years before they became readily apparent in the data residuals between Jason-1 and TOPEX during the cal/val phase.

More importantly, this type of cross calibration has proved invaluable for linking the SSH time series from two different satellites by determining a relative bias. Chambers et al. (2003a) pointed out that local relative biases between Jason-1 and T/P computed with coincident data differ by several cm from the global mean, with slightly larger differences in coastal waters and the Mediterranean Sea, where the largest number of absolute calibration sites is located, and in areas of high SWH due to mean errors in SSB models. Some of the difference has been linked to residual geographic orbit error from the JGM-3 gravity field (Beckley et al., 2004). These results give some explanation for the differences in the bias determined from various *in situ* calibration sites (**Table 2**).

In contrast to the absolute bias computations, the relative biases (Jason relative to TOPEX) estimated from the global SSH residuals during the cal/val phase are far more consistent (**Table 2**). More significantly, there is nearly a 4 cm systematic bias between the mean of the point calibrations and the 'global' calibrations. If one considers the global calibration to be the truth for linking the altimeter data from two different missions to compute a long record of GMSL variation, then relying on an estimate from a single site could be in error by up to 4 cm. While this is not a problem where missions overlap (like T/P and Jason-1), it could be disastrous if there is a long gap between altimeters.

The Jason-2/OSTM mission was also flown in a cal/val phase with Jason-1 for 200 days, on a formation with a temporal separation of only 55 s. The Jason-2/OSTM SSH measurements were initially biased relative to those from Jason-1 by about 7.5 cm based on the global analysis (Ablain et al., 2010) or 8.4 cm based on the calibration at the Harvest platform (Haines et al., 2010). This was surprising considering the Jason-2 altimeter is a clone of the Jason-1 altimeter. However, a majority of the bias has been subsequently traced to a truncated pulse repetition frequency in both altimeters and a difference in the parameterization sets for the Jason-1 and Jason-2 Ku-band ranges (Ablain et al., 2010), which will cause a 9.5 cm bias. The remaining 1–2 cm bias is currently unexplained. A global mapping of the relative bias between Jason-2/OSTM and Jason-1 SSH indicates significant regional patterns, with deviations of more than 1 cm and a strong hemispherical pattern that may be related to orbit error (Ablain et al., 2010).

In summary, the results of the Jason-1 and Jason-2/OSTM cal/val phases have pointed out several important things. First, residual analysis can detect very small errors that may have gone undetected using other methods. This has proved especially true for calibration of the water vapor radiometers. Second, the idea of a global bias in SSH measurements – meaning that the bias is the same everywhere – is a convenient fiction. The SSH measurement depends on a number of measurements and corrections, and a lot of the errors in these constituents can have biases that differ depending on where the measurement is made – that is, SSB errors that are dependent on the mean SWH state or wet troposphere errors that may be consistently larger in regions where there is more water vapor like coastal areas. This means that the global average bias can differ significantly from the local bias. Even considering only the longest record from Harvest (Haines et al., 2003, 2010), the difference between a single local bias estimate and the global average can differ by nearly 3 cm. This is not a trivial difference for continuing long records of GMSL change, which have variations of the order of less than 1 cm and trends of only 2–3 mm year^{-1}.

3.05.5 Applications to Geophysics

3.05.5.1 MSS, Gravity, and Bathymetry

For hundreds of years, the ill-defined term 'mean sea level' has been used as a geodetic reference. At best, mean SL is based on a benchmark near a tide gauge that represents a time average of the local SSH. At worst, the reference is extrapolated between tide gauges or defined by some arbitrary barometric pressure value. Even if one could measure the mean SL around coastlines with millions of tide gauges, such a reference will not be constant over time, since SL is rising for a variety of reasons (Section 3.05.5.3). In addition, the land that tide gauges are

affixed to may also be lifting or subsiding, due to either elastic rebound from the ice loading during the last glaciation, volcanic activity, or pumping of water and oil from underground reservoirs. In our modern age, with satellite altimeters and GPS measuring heights to a few cm or less, a more definitive, stable reference is needed.

Once such reference is the geoid, which is a mathematical surface that represents a specific equipotential surface related to the Earth's gravitational and rotational potential (e.g., Tapley and Kim, 2001). Modern geoids are based on global Earth gravity models determined from satellite tracking data (e.g., Nerem et al., 1994a; Tapley et al., 1996), most recently on data from the GRACE (Tapley et al., 2004b) or the Gravity field and steady-state Ocean Circulation Explorer (GOCE) (Rummel and Gruber, 2010). Geoids will differ depending on what gravity model is used to determine the gravitational potential, as well as what equipotential value is assumed. In recent years, an equipotential determined from an altimetric mean SSH is used.

To first order, the geoid is approximated by an ellipsoid of revolution, defined by an equatorial and polar radius (R_E and R_P) or an equatorial radius and flattening coefficient, $f=((R_E-R_P)/R_E)$. There are many reference ellipsoids in the literature; the one most used in altimetry was defined early in the T/P mission to represent a surface with a zero mean to the measured SSH (e.g., Tapley et al., 1994a). It has values of $R_E=6\,378\,136.3$ m and $f=1/298.257$. The geoid still has significant normal deviations above or below the ellipsoid, of order 100 m. In the literature, these geoid undulations are often referred to as *geoid height* and should not be confused with the full geoid (which includes the portion represented by the ellipsoid).

Ninety-nine percent or more of the signal in the SSH measured by an altimeter with respect to the ellipsoid is the geoid height. The next largest quasipermanent signal is related to the mean ocean geostrophic circulation, which is balanced by gradients in the mean dynamic ocean topography (DOT) (e.g., Tapley et al., 2003; Wunsch and Gaposchkin, 1980). The DOT is the difference between the measured SSH and the geoid and has typical variations of ± 2 m. **Figure 11** shows the mean SSH and geoid along a particular T/P pass, along with the difference. Note that the geoid pictured does not model gravity variations with wavelengths smaller than 100 km wavelength, even though these signals are in the mean SSH. These residual geoid variations appear in the raw DOT measurement. **Figure 12** shows the global DOT filtered in order to suppress these residual geoid signals. The topography slopes result in horizontal pressure gradients that are equally balanced by the Coriolis force caused by ocean currents. This causes a difference in the relationship between the topography gradient and the current in the Northern Hemisphere versus the Southern Hemisphere. The topography is higher to the right of a current and lower to the left in the Northern Hemisphere, while the direction of the slope is reversed in the Southern Hemisphere, with higher topography to the left of the current. One has to be careful about how the permanent lunar and solar tidal effects are represented in the geoid if it is combined with altimetry (Rapp, 1989). The altimeter measures the SSH including the

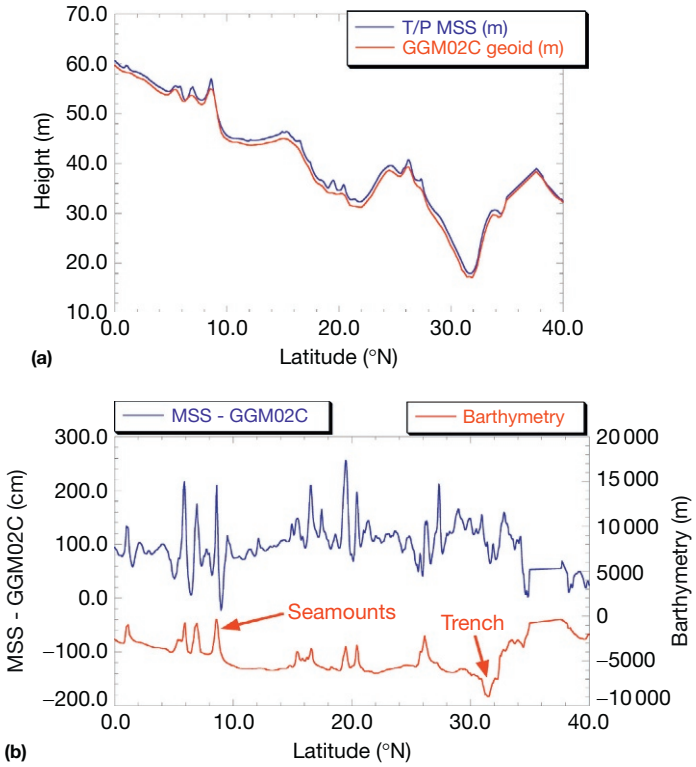


Figure 11 (Top) Average SSH along a T/P pass in the Western Pacific (blue), along with the GGM02C geoid (red) (Tapley et al., 2005). (Bottom) The difference (blue) shows the long-wavelength dynamic topography and higher-resolution geoid that is not modeled. The corresponding bathymetry is shown in red.

mean permanent tide. The geoid can be represented with or without the permanent tide, and the difference is 20 cm over a wavelength of 20 000 km.

The remaining signals in the SSH are caused by time-variable ocean signals such as tides, planetary waves, eddies, fluctuations in the circulation, and exchanges of heat with the atmosphere and water mass with the atmosphere and land. For many oceanographic applications, it is the time-variable component of SSH that is of most interest, and one is most concerned with accurately removing the mean SSH (or MSS) from the data. This is not trivial for two reasons. First, the altimeter does not exactly sample the same location every repeat cycle, although the subsatellite points fall roughly within a region of 7 km along the ground track and ± 1 km on either side of the nominal track at the equator (**Figure 13**). Second, the MSS (geoid) is not constant within this area, or bin, but can have fairly steep gradients in both the along-track and cross track directions. Brenner et al. (1990) estimated that the average gradient over the ocean was about 2 cm km^{-1}, which means that the difference in MSS from one end of the sampling region

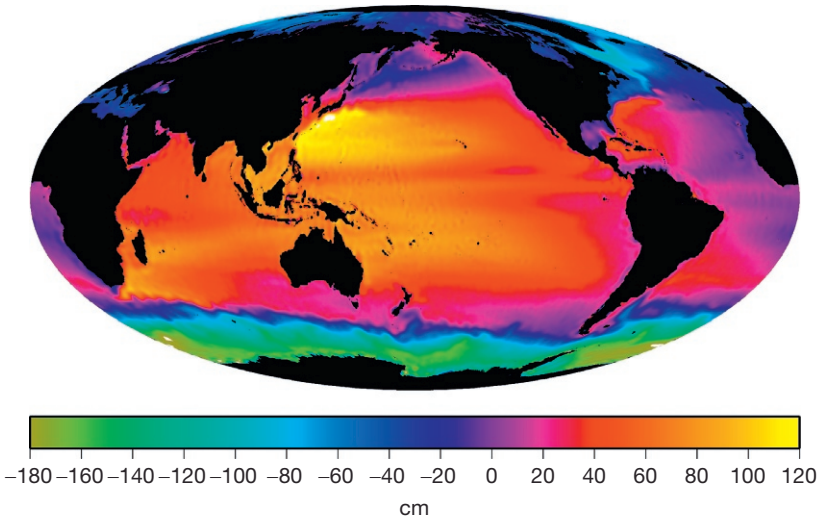


Figure 12 Global DOT based on the Danish National Space Center MSS08 (Andersen and Knudsen, 2009) and the EGM08 geoid (Pavlis et al., 2012), filtered to suppress variations with wavelengths smaller than 111 km.

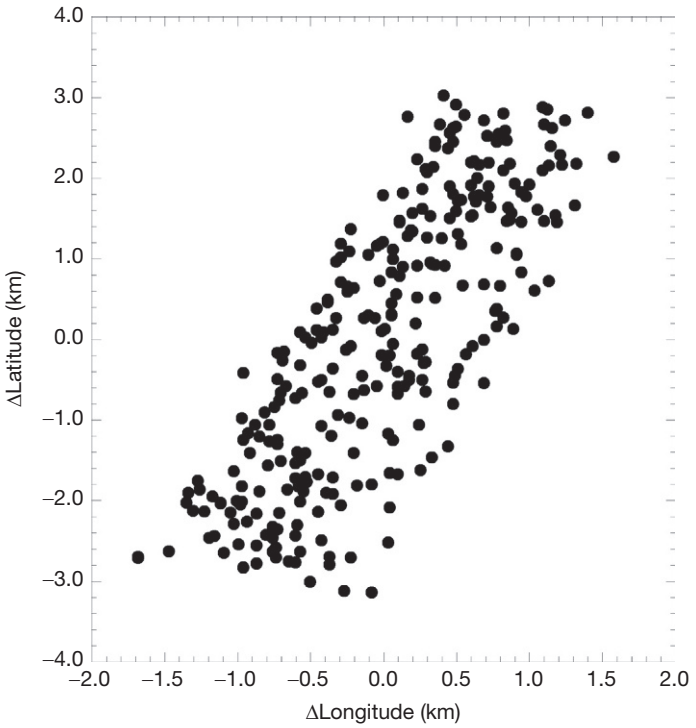


Figure 13 Subsatellite points in one T/P bin for the entire mission.

to the other could be as much as 14 cm. If one simply averaged the SSH points in the region and removed the mean from the data to compute the SSHAs without accounting for the gradients, the geoid slope would appear to be a time-variable signal since the altimeter samples the region in time (e.g., Chambers et al., 1998b). To make matters worse, the geoid slopes are even larger near trenches or seamounts, with values up to 20 cm km^{-1} (Brenner et al., 1990).

One can estimate the along-track gradients by averaging SSH along each ground track to create a MSS profile. This will give the average gradient over ~6–7 km (determined by the sampling rate), which one can then use to correct for the along-track error. However, this approach will not determine the cross track geoid gradient, which could account for up to 4 cm of erroneous SSH variations over most of the ocean, but up to 40 cm near trenches. To account for this, MSS models have been developed based on gridding the MSS profiles from multiple altimeter missions. Then, one can determine and correct for the geoid gradients from these models. Many such models have been produced over the years, differing mainly by altimeter data used, time span of the data, and method used to map the data. Tapley and Kim (2001) discussed the principle methods that have been utilized for mapping the data. The earliest models were based on combinations of Seasat and GEOS-3 data (e.g., Marsh and Martin, 1982; Marsh et al., 1992; Rapp, 1986). The Marsh et al. (1992) model had a spatial sampling of 1/8° (~15 km at the equator), which was only possible because GEOS-3 had a much denser ground track than Seasat. Data from as far away as 100 km from the grid center were used in the gridding process, albeit downweighted as a function of distance. Later, data from the Geosat exact-repeat mission were included (Basic and Rapp, 1992). This did not significantly improve the resolution because it was in the same ground track as Seasat, although errors along the Seasat–Geosat ground track were reduced because of more averaging. The largest improvement in MSS models came when T/P, ERS-1, ERS-2, and Geosat geodetic mission data were included. By the beginning of the Jason-1 mission, several MSS models with resolutions of 1/30° (<4 km at the equator) were available (Andersen and Knudsen, 1998; Cazenave et al., 1996; Tapley and Kim, 2001; Wang, 2000). MSS models have continued to be updated with longer time spans of the T/P, ERS, Jason-1, and Envisat along-track data, as well as data from T/P and Jason-1 on their interleaved ground tracks. Several new MSS models have been released since 2010, including the one constructed at the Danish National Space Center (Andersen and Knudsen, 2009), as well as that of P. Schaeffer and colleagues at CLS (Collecte Localisation Satellites) and CNES in France. The latter model is now included on Jason-2/OSTM GDRs.

Although all modern global MSS models have formal errors of a few cm or less, one has to be careful not to interpret this as the absolute error in the MSS height. One significant problem in MSS mapping is that systematic biases in the SSH measurement will go directly into the MSS and these are not included in the formal error statistics. These errors are caused by residual biases in the range measurement, geographically correlated orbit errors, SSB error, and aliased solar tide error for the Sun-synchronous altimeters. Another source of errors is the use of one-time profiles from the Geosat geodetic mission and ERS-1 168-day repeat mission. These profiles have not averaged out the ocean variability such as mesoscale eddies, which can cause deviations of up to a meter or more. Modelers have attempted to reduce such errors by referencing the MSS profiles from other altimeters to those of T/P (e.g., Tapley and Kim, 2001), but this will not remove systematic errors in T/P, nor will it remove all of the systematic error from the other altimeters at the shortest wavelengths. If one differences two modern MSS models (**Figure 14**), one observes differences of several tens of centimeters in some areas. Although the larger-scale differences (such as in the Western Pacific) are due to different averaging intervals, a significant portion can only be attributed to errors in one or both of the MSS models, for example, in the Arctic.

Another approach to estimating geoid gradients directly along the ground track of a particular satellite was introduced by Chambers et al. (1998b). They estimated the parameters to a two-dimensional plane for a series of bins defined along a nominal ground track based on several years of SSH

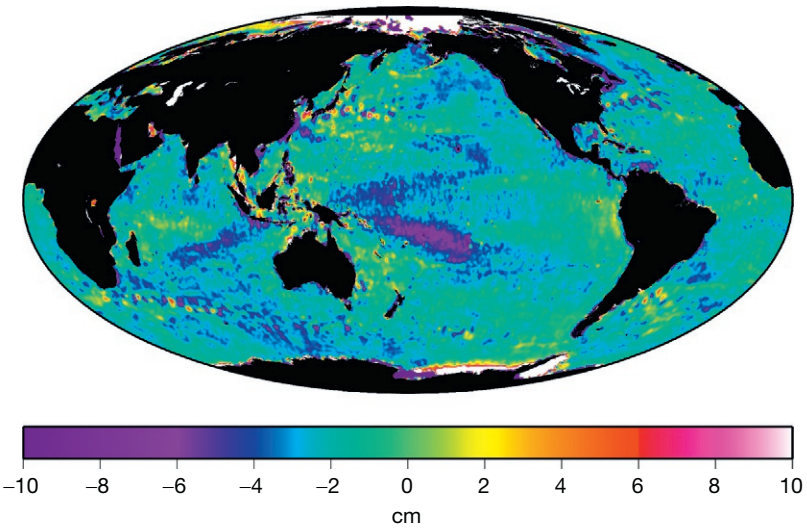


Figure 14 Difference between GSFCMSS00 (Wang, 2000) and CSRMSS98 (Tapley and Kim, 2001), averaged in 0.25 grids.

measurements for the T/P mission. They demonstrated that this local along-track MSS model reduced the variance of SSHAs better than available gridded MSS models because it was tuned to the particular satellite and had the highest possible resolution and that it could be updated quickly as more data were available. One problem with the approach was the aliasing of true SL variability into the estimation of the gradients, although later refinements have reduced the problem by either including harmonic parameters to account for the signals (Chambers, 2002) or first removing gridded maps of long-wave SL variability (Dorandeu et al., 2003).

The MSS can be used as a direct measure of the geoid height with an error equivalent to about 1% of the signal, arising from the nongravitational dynamic topography signal. Often, a topography based on historical measurements (e.g., Levitus, 1982) has been subtracted from the MSS model, and the residual is used as a measure of geoid height. However, the ocean topography model is not perfect, nor will it represent the smallest-scale gradients related to the circulation, such as those near intense jets like the Gulf Stream and Kuroshio. More importantly, systematic errors caused by the altimeter measurements in the MSS will remain.

Such proxies of geoid height over the ocean from altimetry are still vital in the determination of global geoid models in combination with tracking data from satellites (e.g., Lemoine et al., 1998; Pavlis et al., 2012; Tapley and Kim, 2001), since they are the only direct measure of short-wavelength (7–500 km) variations of ocean gravity. There have been problems in previous global models, though, arising from the difficulty in separating gravity from oceanographic currents in the MSS. Until recently, satellite tracking data alone have not been able to accurately resolve the geoid at the wavelengths where this is important (e.g., Tapley et al., 1994b). Tapley et al. (2003) demonstrated that the long-wavelength portion of the EGM96 geoid (Lemoine et al., 1998) actually contained significant oceanographic signals from the combination of the altimetry and satellite tracking data in the solution, which meant it could not be combined with altimetry to determine the DOT. This has become less of a problem when tracking data from the GRACE are utilized in the combination (Tapley et al., 2003), as the GRACE data alone can be used to determine the long-wavelength geoid to an accuracy of a few cm over wavelengths of several hundred kilometers (Tapley et al., 2005). Subsequent improvements in geoids have led to even smaller uncertainties over smaller spatial scales (approaching 100 km), especially using early releases of GRACE/GOCE combination geoids (Bingham et al., 2011; Knudsen et al., 2011).

Other methods have been used to isolate only the short-wavelength gravity information from the altimetry. Haxby et al. (1983) first showed how to differentiate a gridded MSS in order to obtain a map of gravity anomalies. One benefit of differentiating the MSS measurement is the reduction in some systematic errors, especially those with very long wavelength, although it may not eliminate all errors where data from multiple altimeter have been used. Because of this, another method to map gravity anomalies from altimetry was developed (e.g., Sandwell, 1992; Sandwell and Smith, 1997). Instead of differentiating a pregridded MSS model, SSH gradients along each satellite pass were computed and averaged and then converted to gravity anomalies and gridded. By doing this, more of the systematic long-wavelength errors for each particular satellite were eliminated before the data were mapped.

Much has been learned about the geophysics of the seafloor from such altimetric geoid or gravity anomaly maps. The primary cause of change in the geoid over short spatial scales is variations in the seafloor topography. Thus, by studying the altimetric geoid or gravity anomaly maps, one can infer information about seamounts, trenches, spreading ridges, or subduction zones (e.g., Cazenave and Royer, 2001). The information can also be used infer the seafloor topography, or bathymetry (e.g., Sandwell and Smith, 2001; Smith and Sandwell, 1994), although there is one important limitation to this. The gravity to topography transfer function is related to the ocean depth and causes attenuation of the signal in SSH (Sandwell and Smith, 2001). For example, if two topography features are very close together (on the order of the ocean depth or less), they will appear in the SSH as a single feature. Very sharp changes in the topography are also spread out in the SSH. Thus, a step function in the bathymetry would appear as a more gradual slope in the SSH. Still, the gravitational effect of the bathymetry is quite evident in the SSH (**Figure 11**).

3.05.5.2 SL Change

One of the major benefits of satellite altimetry is the ability to measure SL variations in the deep ocean away from coasts. Although local SL change has been measured for more than a century in some areas using tide gauges, it was not possible to measure in the open ocean before the advent of satellite altimeters. Additionally, tide gauges measure SL relative to the local land motion, which may be rising or falling due to postglacial rebound (PGR), volcanic activity, or pumping water and oil from underground reservoirs, and the tide gauges are not tied to a common reference frame. A satellite altimeter, on the other hand, is tied to a consistent reference frame and so measures the total (not relative) SL change to the best of our capability (e.g., Nerem et al., 1998).

Although tide gauges with very long records show a reasonably consistent rate of SL change over the last 100 years of about 1.8 mm year^{-1} when corrected for PGR (Douglas, 1991, 2001; Holgate, 2007; Merrifield et al., 2009), there has been some question whether this truly represents the rate of GMSL, since all but a handful of the two dozen gauges are on the east coast of North America and the west coast of Europe. Because of its nearly global sampling, satellite altimetry was expected to resolve some of these questions.

Computing GMSL from altimetry is relatively straightforward, and most analyses use a procedure similar to that described in more detail by Nerem (1995) with only a few modifications. Essentially, the SSH measurements along each ground track pass are reduced to SSHAs about the mean SSH using either a mean profile, a global MSS model, or satellite-specific along-track MSS model (**Section 3.05.5.1**). The SSHAs for each repeat cycle are then averaged, accounting for the fact that there are more observations in the high latitudes because of the ground track spacing (Nerem, 1995). From this, one obtains a point representing the GMSL for each repeat period. With a long enough time series, the rate of GMSL change can be obtained. Note that although GMSL is often used in the

literature, it is not meant to suggest that the entire ocean is rising or falling by this amount. GMSL is really just an average of local SL variations, which are generally higher than or lower than the global average. However, GMSL is often used to study small, subtle variations and trends related to climate change that may or may not be readily apparent in the local measurements.

The Seasat mission did not last long enough to make regular measurements of GMSL, although a determination was made using a month of data (Born et al., 1986). At best, its data represent a single measurement of GMSL in 1978 with a fairly large error bar due to the limitations in a single-site calibration, especially one without a large number of observations over time (**Section 3.05.4.1**). The Geosat exact-repeat mission lasted nearly 3 years from 1986 to 1989, but the GMSL measurement was corrupted by errors in the orbit and range corrections (Wagner and Cheney, 1992). The determination of the GMSL rate from Geosat ranged from 0 mm year^{-1} (Tapley et al., 1992) to −12 mm year^{-1} (Wagner and Cheney, 1992), and because the altimeter was never calibrated like Seasat, it was difficult to tie the two measurements together.

Although ERS-1 was launched over a year before T/P, it too suffered from large errors in the orbit, as well as problems in the instrument and atmospheric corrections, which limited the ability to determine GMSL. It was not until several years of reprocessing and corrections had been made that GMSL could be determined from ERS-1 data (e.g., Anzenhofer and Gruber, 1998), and by this point, a much longer and more accurate time series was being determined regularly from T/P (e.g., Nerem et al., 1997). It is only after considerable effort of reprocessing and reprocessing that observations from ERS-1 and ERS-2 have been able to be computed and integrated into the GMSL record (Leuliette and Scharroo, 2010).

The earliest estimates of GMSL rate from T/P (Minster et al., 1995; Nerem, 1995) were unfortunately corrupted by the error in the oscillator correction that caused the estimate to be nearly 7 mm year^{-1} too high (Nerem et al., 1997; **Section 3.05.4.2**). Intermediate results were affected by a smaller drift in the TMR brightness temperatures (**Section 3.05.2.3**), but this was soon corrected and reasonable estimates of GMSL from T/P were being calculated (e.g., Nerem et al., 1999). Finally, the change from TOPEX_A to TOPEX_B caused another bias in T/P GMSL of nearly 1 cm after February 1999 that was corrected when the SSB model was updated (Chambers et al., 2003b; **Section 3.05.2.4**). **Figure 15** shows the latest time series of GMSL from corrected T/P, Jason-1, and Jason-2/OSTM data. The trend over the 19-year period is 3 mm year^{-1}. This is confirmed by multiple groups utilizing the same data (Ablain et al., 2009; Beckley et al., 2010; Leuliette and Scharroo, 2010; Nerem et al., 2010). Leuliette et al. (2004) estimated that the error in this determination is ±0.4 mm year^{-1}, based on a combination of the formal error in the fit and the accuracy of calibration relative to tide gauges.

Note that the RMS for the difference between the T/P and Jason-1 measurements during the 210 day overlap is 4 mm, which suggests an error in any 10-day estimate of less than 3 mm RMS. This indicates that the variations about the long-term linear trend are not noise, but represent real GMSL variations at shorter periods. The seasonal variation of approximately 4 mm amplitude has been shown to be caused by a combination of seasonal transport of water mass between the hydrosphere and ocean and annual changes in the ocean heating (Chen et al., 1998; Minster et al., 1999). The mass variation has an amplitude of 8–10 mm but is 180 out of phase with a smaller (4 mm) ocean heating signal (see also Section 65.3.4 of this volume). Although this is a very small signal, it is detectable in the altimeter measurements.

Interannual fluctuations even larger than the mean seasonal variation are apparent in **Figure 15**, especially in during 1997/1998 and in 2011. GMSL rose steadily by nearly 25 mm before peaking at the end of 1997 and then falling again in the subsequent year. This was linked to the 1997/1998 El Niño event by Nerem et al. (1999), who suggested that SL change was related to a reduction in the global mean outgoing longwave radiation at the same time, causing the upper ocean to store more heat during this period. Additionally, they demonstrated that a model that assimilated *in situ* temperature data also reproduced the signal.

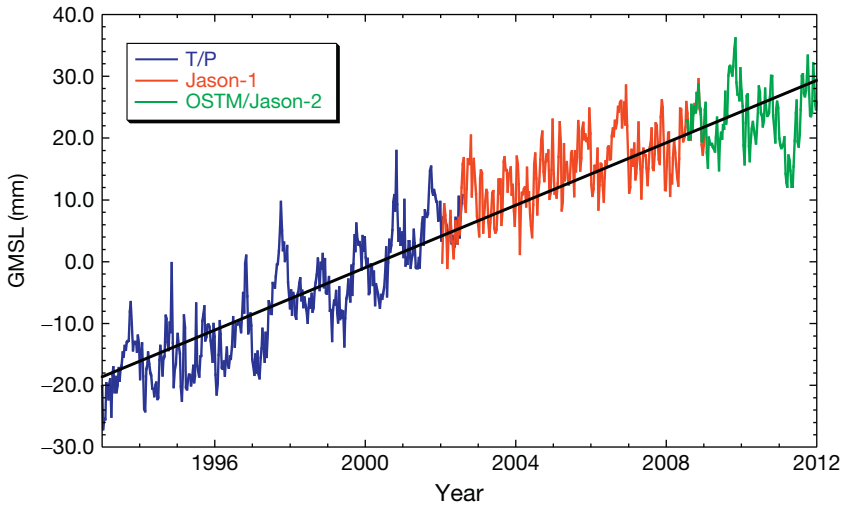


Figure 15 GMSL from corrected T/P (blue), Jason-1 (red), and OSTM/Jason-2 (green) data from January 1993 to April 2012 at 10-day intervals. The T/P data have been corrected for the oscillator error (**Section 3.05.4.2**), the drift in TMR (**Section 3.05.2.3**), and the different TOPEX_A and TOPEX_B SSB models (**Section 3.05.2.4**), while the Jason-1 and Jason-2 data have had a relative bias applied (**Section 3.05.4.1**) and have been corrected for known biases in the wet path delays (**Section 3.05.2.3**). The long-term linear trend is shown with a thick black line.

However, Willis et al. (2004) utilized a mapping of available *in situ* temperature data and found no evidence of a significant increase in upper ocean heat content during the 1997/1998 El Niño. They suggested the most likely source was a temporary gain in ocean mass due to interannual changes in the water cycling between the continents and oceans. The drop in SL in 2011 was caused by precipitation and water cycling changes over land associated with a large La Niña event (Boening et al., 2012). Thus, the large change in 1997/1998 is quite likely also related to fluctuations in the water cycle, not heating.

The discovery that GMSL is affected to such a large amount by interannual variations in like El Niño and La Niña has led to some concern that part of the difference between the long-term rates measured by tide gauges over 100 years (~1.8 mm year^{-1}) and altimetry over <20 years (~3 mm year^{-1}) may be due to interannual to decadal variations in GMSL. Nerem et al. (1999) attempted to estimate the effect of El Niño/La Niña using the Southern Oscillation Index scaled to GMSL and found that with 10 years of continuous data, the error introduced from El Niño-period variability would be less than 0.5 mm year^{-1}. However, more recent studies of long tide gauge records indicate strong multidecadal fluctuations with a period near 60 years in numerous gauges (Chambers et al., 2012). It is not yet clear whether these longer-period fluctuations are also reflected in GMSL.

Altimetry also gives a unique perspective on change in local rates of SL, which is impossible from tide gauge data without significant extrapolation. Local rates (**Figure 16**) determined over the same period of time as the GMSL curve in **Figure 15** show that there are only a few locations where the local rate is the same as the GMSL rate. In most places, it is either significantly higher (as much as two to three times larger) or actually falling. We know that the patterns change significantly during El Niño years in the tropics when there are large local changes in the heat storage; however, the pattern of higher SL trends in the Western Pacific has persisted over the last decade regardless of El Niño or La Niña. This pattern has recently been shown to have started shortly before the launch of T/P (Merrifield, 2011) and is caused by an intensification of the trade winds (Merrifield and Maltrud, 2011) that may be related to the PDO (Merrifield et al., 2012).

In addition to studying the patterns of SL rate, the authors have used maps of SL anomalies determined from altimeters to study other patterns of SL variability, using empirical orthogonal functions (EOFs) (e.g., Chambers et al., 2002; Hendricks et al., 1996). EOF analysis is used to find correlated patterns in a data set, where the patterns are represented by a spatial mode (EOFs) and an associated time series (principal components). The modes are sorted by the amount of variance of the total signal explained. Thus, the first mode will explain the largest percentage of variability and the last modes will explain the least. Chambers et al. (2002) used such spatial patterns estimated from T/P data in order to reconstruct global maps of SL variability from tide gauge data, using an idea first used with sea surface temperature data (Smith et al., 1996). The idea is to estimate the associated time series from the tide gauge data using the altimeter patterns as a basis function. Chambers et al. (2002) computed only the interannual variations after removing all long-term trends from tide gauge data and then derived a GMSL time series from the reconstructed fields back to 1950. They found significant El Niño variability throughout the time period and verified the results in the 1980s using Geosat data and a proxy for SL derived from sea surface temperature records. Church et al. (2004) recently extended this idea to better determine long-term GMSL rates from tide gauges by using the EOF patterns from T/P to model interannual variations. This method has continued to be used to reconstruct GMSL back to before 1900, using EOFs from longer satellite altimeter records (Church and White, 2011; Ray and Douglas, 2011). As even longer multidecadal altimeter records become available, spatial patterns associated with decadal variations will become detectable, and this will allow for better understanding of the low-frequency GMSL variations.

3.05.6 Conclusions and Future Prospects

Satellite radar altimetry has become a powerful tool for measuring both geophysical and oceanographic signals with high precision. The accuracy of the SSH measurement is now approaching that of a tide gauge, but no comparable instrument can match the satellite altimeter for the sheer density of

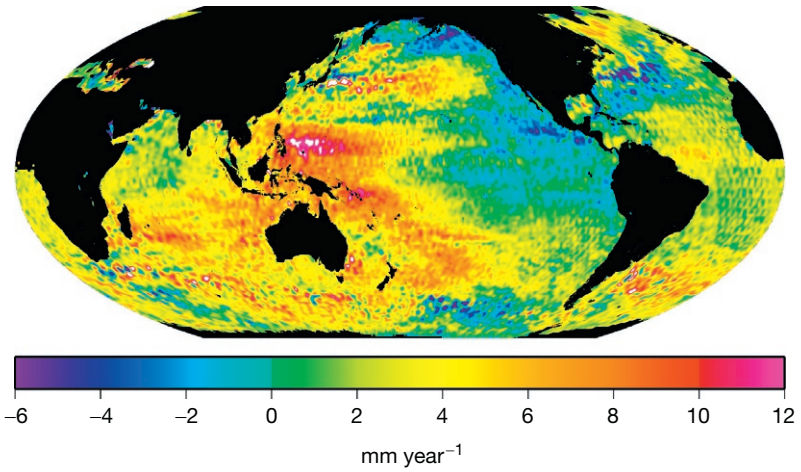


Figure 16 Map of local SL rates determined from T/P, Jason-1, and OSTM/Jason-2 data from January 1993 to December 2011. Data were mapped using optimal interpolation and a covariance function as determined from the full record of SSH. All data were corrected as for **Figure 15**.

measurements in such a short period. For the first time, we are truly obtaining global measurements of SL on a regular basis. Data from current missions are already being used in operational data-assimilating models for tracking ocean eddies and current variations and predicting El Niño and hurricane intensification. Future radar altimeter missions using instruments similar to the ones discussed here are planned. These will not significantly advance the capability of the measurement, but merely continue the length of the data records. Several new types of altimeter systems that have significantly different resolution or precision over a radar altimeter have been proposed, tested, or flown in experimental modes. In concluding this chapter, we would like to briefly describe some of these and comment on their possible use for advancing the science of geophysics.

3.05.6.1 Laser Altimetry

Two-laser altimeters have already been operated from satellite platforms. The Mars Orbiter Laser Altimeter (MOLA) was launched in 1996 aboard the Mars Global Surveyor mission and conducted a multiyear mapping mission of the red planet beginning in early 1999 (Smith et al., 2001). An earlier version of the MOLA instrument had been on the Mars Observer mission that never reached orbit. The Ice, Cloud, and Land Elevation Satellite (ICESat) was launched into the Earth's orbit in January 2003 and carried the Geoscience Laser Altimeter System (GLAS), with a primary mission to measure changes in the Earth's ice sheet elevation over time (Schutz et al., 2005; Zwally et al., 2002). Although radar altimeter data have been used to measure ice sheets for some time (e.g., Zwally and Brenner, 2001), the accuracy of the radar measurement over the ice is much worse than over the ocean, because ice sheet elevation changes so much over the footprint of the radar pulse.

The footprint of a laser is significantly smaller than that of a microwave pulse, typically 70 m compared to the effective footprint for a radar altimeter of order 10 km. This allows for the along-track separation of unique data points to be much closer, about 200 m or less compared to 7 km for the radar, and hence, the laser altimeter can detect steep changes in the surface elevation with higher accuracy than a radar altimeter. However, the quality that makes the laser altimeter so good over ice sheets degrades its performance over the ocean compared to a radar altimeter. Because the along-track separation and footprint size of the laser altimeter are smaller than the wavelength of typical waves, the laser altimeter measures the wave surface of the ocean, not the mean surface like a radar altimeter. However, by averaging over multiple footprints to reduce wave effects, one can compute a SSH that is similar to that from radar altimeter, except in the areas with highest waves (Urban and Schutz, 2005). Although unlikely to replace radar altimetry for measuring the time-variable sea surface, laser altimetry may be useful to determine a much higher resolution of the MSS than is possible with any radar altimeter system (Zwally et al., 2002). This however has not been possible to demonstrate with ICESat data due to issues involving reduced instrument operation times than originally planned.

The first laser on ICESat failed after only 36 days of operation. There were three separate lasers on ICESat, and each was designed to operate for at least 1 year in order to meet a mission lifetime of 3 years. After an independent review (Kichak, 2003), it was theorized that the most likely cause was a short in the diode array and that the same failure could occur in the other two lasers. It was recommended that a reduced science mode be planned, operating the lasers at a lower temperature and not continuously in order to extend the lifetime as long as possible. GLAS-2 operated for three periods during 2003 and 2004 before failing, while GLAS-3 operated in the reduced observing mode until October 2009. The typical operating period was for a month with 3–4-month periods offline, typically focusing on periods of peak ice loss on Greenland and Antarctica. 15 full month-long campaigns were completed.

Although ICESat did not meet all of its science goals, much like Seasat, it did demonstrate the capability of the laser altimeter for Earth geoscience applications, especially measuring the rapid changes in ice sheet thickness of the Greenland and Antarctic ice sheets (e.g., Gunter et al., 2009; Howat et al., 2008; Luthcke et al., 2005; Pritchard et al., 2009; Sorensen et al., 2011). An ICESat-2 mission is currently being planned by NASA, with a launch date of 2016–17. Besides improving the problems in the lasers that led to premature failure on ICESat, ICESat-2 will use a micropulse multibeam approach, to allow for denser cross track sampling in order to better resolve surface slope for each pass, without needing multiple passes to resolve.

3.05.6.2 Delay-Doppler Altimeter

A delay-Doppler (DD) radar altimeter measures the phase change of the return signal instead of the travel time, using techniques developed for synthetic aperture radars. This leads to improved precision by a factor of 2 or more at the sampling rate of a standard radar altimeter (Raney, 1998), which means that a DD altimeter can make samples more frequently than a radar altimeter with comparable precision. The DD altimeter also has a smaller effective footprint than a standard radar altimeter, which means it will be more accurate in places with steep changes in the topography, such as over ice sheets.

John Hopkins University Applied Physics Laboratory (APL) has built, tested, and demonstrated DD altimeters on aircraft for the last decade (APL, 2005). The first space-based DD altimeter was installed on the CryoSat satellite to measure changes in the Earth's ice sheets. Unfortunately, CryoSat crashed into the Arctic Ocean due to a launch failure on 8 October 2005. A CryoSat 2 mission was approved less than 6 months later, and several improvements were made to the spacecraft/instrument in development, including adding a redundant backup altimeter (ESA, 2007). CryoSat 2 was launched on 8 April 2010. No papers have been published as of this writing on scientific results from CryoSat data over the ice sheets, but preliminary analysis presented at conferences suggests the observations are consistent with expectations (e.g., Shepherd et al., 2010). A preliminary study examining the CryoSat data over the ocean indicates that although it was not designed as an ocean-observing mission, using CryoSat SSH measurements in combination with Envisat and Jason-1 and Jason-2 SSH measurements improves mapping of mesoscale eddies (Dibarboure et al., 2012) and also agrees well with

coincident altimetry from Envisat on the broadest scales (Labroue et al., 2012).

3.05.6.3 Wide Swath Altimeter

Another technique has been proposed to measure the SSH with an even higher spatial density than is even possible with a DD altimeter, based on a radar interferometer (Rodriguez and Martin, 1992; Rodriguez and Pollard, 2001), and known as a wide swath altimeter (WSA). The details of the instrument are beyond the scope of this document, but the basic concept is that by placing two interferometric antennas on the ends of several meter booms attached to a satellite bus with a nadir radar altimeter, one could obtain a swath of SSH measurements several hundred km wide, instead of the 2 km wide swath with a typical radar altimeter (Rodriguez and Pollard, 2001). The nadir instrument would be used for estimation of ionosphere and tropospheric path delays, as well as calibration of the interferometer altimeter data.

Such an instrument was being seriously considered for inclusion in the follow-on mission to Jason-1 – the OSTM. It was estimated that the WSA on OSTM would provide 10 day SSHA maps with approximately 15 km resolution, with only a few gaps between crossovers (Rodriguez and Pollard, 2001). Although the accuracy of the off-nadir points was predicted to not be as good as the nadir points, the difference was not too great (Rodriguez and Pollard, 2001). Unfortunately, even though work on the WSA had progressed significantly beyond a simple conceptual model, budget limitations and concerns about a slippage of the planned launch date forced its removal from the final OSTM design.

The idea of a spaceborne WSA was subsequently embraced by the hydrology community (e.g., Alsdorf et al., 2007), who proposed a mission to fly a WSA to better measure small bodies of surface water that were impossible to detect with current radar altimeters. The two concepts have since been combined into a single mission, the Surface Water Ocean Topography (SWOT) mission (Fu et al., 2012), which is being developed by NASA and CNES. Although the mission has not been formally approved as of this writing, it is expected that SWOT or a similar mission will be flown within the next decade and help us improve our understanding of sub-mesoscale processes of the ocean, as well as small-scale water elevation changes over land.

References

Ablain M, Cazenave A, Guinehut S, and Valladeau G (2009) A new assessment of global mean sea level from altimeters highlights a reduction of global slope from 2005 to 2008 in agreement with in-situ measurements. *Ocean Science* 5: 193–201.

Ablain M, Philipps S, Picot N, and Bronner E (2010) Jason-2 global statistical assessment and cross-calibration with Jason-1. *Marine Geodesy* 33: 162–185.

Alsdorf DE, Rodriguez E, and Lettenmaier D (2007) Measuring surface water from space. *Reviews of Geophysics* 45: RG2002. http://dx.doi.org/10.1029/2006RG000197.

Andersen OB and Knudsen P (1997) Multi-satellite ocean tide modeling – the K_1 constituent. *Progress in Oceanography* 40: 197–216.

Andersen OB and Knudsen P (1998) Global marine gravity field from the ERS-1 and Geosat geodetic mission altimetry. *Journal of Geophysical Research* 103: 8129–8137.

Andersen OB and Knudsen P (2009) DNSC08 mean sea surface and mean dynamic topography models. *Journal of Geophysical Research* 114: C11001. http://dx.doi.org/10.1029/2008JC005179.

Anzenhofer M and Gruber T (1998) Fully reprocessed ERS-1 altimeter data from 1992 to 1995: Feasibility of the detection of long-term sea level change. *Journal of Geophysical Research* 103: 8089–8112.

Applied Physics Laboratory (2005) http://fermi.jhuapl.edu/d2p/. John Hopkins University Applied Physics Laboratory.

Basic T and Rapp RH (1992) *Oceanwide Prediction of Gravity Anomalies and Sea Surface Heights Using GEOS 3, Seasat, and Geosat Altimeter Data and ETOPO5U Bathymetric Data*. Columbus, OH: Ohio State University, DGSS Report 416, 89 pp.

Beckley BD, Zelensky NP, Holmes SA, et al. (2010) Assessment of the Jason-2 extension to the TOPEX/Poseidon, Jason-1 sea-surface height time series for global mean sea level monitoring. *Marine Geodesy* 33: 447–471.

Beckley BD, Zelensky NP, Luthcke SB, and Callahan PS (2004) Towards a seamless transition from TOPEX/Poseidon to Jason – 1. *Marine Geodesy* 27: 373–390.

Bertiger W, Desai SD, Dorsey D, et al. (2010) Sub-centimeter precision orbit determination with GPS for ocean altimetry. *Marine Geodesy* 33(S1): 363–378.

Bilitza D (1997) International reference ionosphere – Status 1995/96. *Advances in Space Research* 20: 1755–1759.

Bingham RJ, Knudsen P, Andersen O, and Pail R (2011) An initial estimate of the North Atlantic steady-state geostrophic circulation from GOCE. *Geophysical Research Letters* 38: L01606. http://dx.doi.org/10.1029/2010GL045633.

Boening C, Willis JK, Landerer FW, Nerem RS, and Fasullo J (2012) The 2011 La Niña: So strong, the oceans fell. *Geophysical Research Letters* 39: L19602. http://dx.doi.org/10.1029/2012GL053055.

Bonnefond P, Exertier P, Laurain O, et al. (2003) Absolute calibration of Jason-1 and TOPEX/Poseidon altimeters in Corsica. *Marine Geodesy* 26: 261–284.

Born GH, Lemoine FG, and Crawford MJ (1988) Geosat ERM—Orbit determination. *Advanced Aeronautical Science* 65: 65–81.

Born GH, Mitchell JL, and Heyler GA (1987) Geosat ERM—Mission design. *Journal of Astronautical Science* 35: 119–134.

Born GH, Parke ME, Axelrad P, et al. (1994) Calibration of the TOPEX altimeter using a GPS buoy. *Journal of Geophysical Research* 99(C12): 24517–24526. http://dx.doi.org/10.1029/94JC00920.

Born GH, Tapley BD, Ries JC, and Stewart RH (1986) Accurate measurement of mean sea level changes by altimetric satellites. *Journal of Geophysical Research* 91: 11775–11782.

Brenner AC, Koblinsky CJ, and Beckley BD (1990) A preliminary estimate of geoid-induced variations in repeat orbit satellite altimeter observations. *Journal of Geophysical Research* 95: 3033–3040.

Brown S, Desai S, Keihm S, and Lu W (2009) Microwave radiometer calibration on decadal time scales using on-earth brightness temperature references: Application to the TOPEX microwave radiometer. *Journal of Atmospheric and Oceanic Technology* 26: 2579–2591. http://dx.doi.org/10.1175/2009JTECHA1305.1.

Brown S, Ruf C, Keihm S, and Kitiyakara K (2004) Jason-1 microwave radiometer performance and on-orbit calibration. *Marine Geodesy* 27: 199–220.

Cao C, Chen R, and Miller L (2011) Monitoring the Jason-2/AMR stability using SNO observations from AMSU on MetOp-A. *Marine Geodesy* 34: 431–446.

Cartwright DE and Ray RD (1990) Oceanic tides from GEOSAT altimetry. *Journal of Geophysical Research* 95: 3069–3090.

Cazenave A and Royer JV (2001) Applications to marine geophysics. In: Fu L-L and Cazanave A (eds.) *Satellite Altimetry and Earth Science*. *International Geophysics Series*, vol. 69, pp. 407–439. San Diego: Academic Press, Chapter 11.

Cazenave A, Schaeffer P, Berge M, Brossier C, Dominh K, and Gennero MC (1996) High-resolution mean sea surface computed with altimeter data of ERS-1 (geodetic mission) and Topex-Poseidon. *Geophysical Journal International* 126: 696–704.

Chambers DP (2002) Effect of sea level variability on the estimation of mean sea surface gradients. *Marine Geodesy* 25: 273–288.

Chambers DP, Hayes SA, Ries JC, and Urban TJ (2003a) New TOPEX sea state bias models and their effect on global mean sea Level. *Journal of Geophysical Research* 108(C10): 3305. http://dx.doi.org/10.1029/2003JC001839.

Chambers DP, Mehlhaff CA, Urban TJ, Fujii D, and Nerem RS (2002) Low frequency variations in global mean sea level: 1950–2000. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JC001089.

Chambers DP, Merrifield MA, and Nerem RS (2012) Is there a 60-year oscillation in global mean sea level? *Geophysical Research Letters* 39: L18607. http://dx.doi.org/10.1029/2012GL052885.

Chambers DP, Ries JC, Shum CK, and Tapley BD (1998a) On the use of tide gauges to calibrate altimeter drift. *Journal of Geophysical Research* 103: 12885–12890.

Chambers DP, Ries JC, and Urban TJ (2003b) Calibration and verification of Jason-1 using global along-track residuals with TOPEX. *Marine Geodesy* 26: 305–318.

Chambers DP, Tapley BD, and Stewart RH (1998b) Reduction of geoid gradient error in ocean variability from satellite altimetry. *Marine Geodesy* 21: 25–39.

Chelton DB (1994) The sea-state bias in altimeter estimates of sea level from collinear analysis of TOPEX data. *Journal of Geophysical Research* 99: 24995–25008.

Chelton DB, Ries JC, Haines BJ, Fu L-L, and Callahan PS (2001) Satellite altimetry. In: Fu L-L and Cazanave A (eds.) *Satellite Altimetry and Earth Science*. *International Geophysics Series*, vol. 69, pp. 1–131. Academic Press, Chapter 1.

Chelton DB and Schlax MG (1993) Spectral characteristics of time-dependent orbit errors in altimeter height measurements. *Journal of Geophysical Research* 98: 12579–12600.

Chelton DB and Schlax MG (1996) Global observations of oceanic Rossby waves. *Science* 272: 234–238.

Chen JL, Wilson CR, Chambers DP, Nerem RS, and Tapley BD (1998) Seasonal global water mass balance and mean sea level variations. *Geophysical Research Letters* 25: 3555–3558.

Cheney RE, Doyle NS, Douglas BC, et al. (1991) *The Complete Geosat Altimeter GDR Handbook*. Rockville, MD: US Department of Commerce, NOAA Manual NOS NGS 7.

Cheney RE and Miller L (1990) Recovery of the sea level signal in the western tropical Pacific from Geosat altimetry. *Journal of Geophysical Research* 95(C3): 2977–2984. http://dx.doi.org/10.1029/JC095iC03p02977.

Cheney R, Miller L, Agreen R, Doyle N, and Lillibridge J (1994) TOPEX/POSEIDON: The 2-cm solution. *Journal of Geophysical Research* 99: 24555–24563.

Choi K-R, Ries JC, and Tapley BD (2004) Jason-1 precision orbit determination by combining SLR and DORIS with GPS tracking data. *Marine Geodesy* 27: 319–332.

Christensen EJ, Haines BJ, Keihm SJ, et al. (1994) Calibration of TOPEX/POSEIDON at platform harvest. *Journal of Geophysical Research* 99: 24465–24485.

Church JA and White NJ (2011) Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics* 32(4–5): 585–602.

Church JA, White NJ, Coleman R, Lambeck K, and Mitrovica JX (2004) Estimates of the regional distribution of sea-level rise over the 1950 to 2000 period. *Journal of Climate* 17: 2609–2625.

Daniell RE, Brown LD, Anderson DN, et al. (1995) Parameterized ionospheric model: A global ionospheric parameterization based on first principles models. *Radio Science* 30: 1499–1510.

Desai SD and Haines BJ (2004) Monitoring measurements from the Jason-1 microwave radiometer and independent validation with GPS. *Marine Geodesy* 27: 221–240.

Dibarboure G, Renaudie C, Pujol M-I, Labroue S, and Picot N (2012) A demonstration of the potential of Cryosat-2 to contribute to mesoscale observation. *Advances in Space Research* 50: 1046–1061. http://dx.doi.org/10.1016/j.asr.2011.07.002.

Dickman SR (1988) Theoretical investigation of the oceanic inverted barometer response. *Journal of Geophysical Research* 93: 14941–14946.

Dorandeu J, Ablain M, Faugere Y, Mertz F, Soussi B, and Vincent P (2004) Jason-1 global statistical evaluation and performance assessment: Calibration and cross-calibration results. *Marine Geodesy* 27: 345–372.

Dorandeu J, Ablain M, and Le Traon P-Y (2003) Reducing cross-track geoid gradient errors around TOPEX/Poseidon and Jason-1 nominal tracks: Application to calculation of sea level anomalies. *Journal of Atmospheric and Oceanic Technology* 20: 1826–1838.

Dorandeu J and Le Traon P-Y (1999) Effects of global mean atmospheric pressure variations on mean sea level changes from TOPEX/Poseidon. *Journal of Atmospheric and Oceanic Technology* 16: 1279–1283.

Douglas BC (1991) Global sea level rise. *Journal of Geophysical Research* 96: 6981–6992.

Douglas BC (2001) Sea level change in the era of the recording tide gauge. In: Douglas BC, Kearney M, and Leatherman S (eds.) *Sea Level Rise: History and Consequences*, pp. 37–64. Academic Press.

Douglas BC, Agreen RW, and Sandwell DT (1984) Observing global ocean circulation with SEASAT altimeter data. *Marine Geodesy* 8: 67–83.

Elfouhaily T, Thompson DR, Chapron B, and Vandemark D (2000) Improved electromagnetic bias theory. *Journal of Geophysical Research* 105: 1299–1310.

Elfouhaily T, Thompson DR, Vandemark D, and Chapron B (1999) Weakly nonlinear theory and sea state bias estimations. *Journal of Geophysical Research* 104: 7641–7647.

Emery WJ, Born GH, Baldwin DG, and Norris CL (1990) Satellite-derived water vapor corrections for Geosat altimetry. *Journal of Geophysical Research* 95: 2953–2964.

ESA (2007) CryoSat mission and data description. ESA Report CS-RP-ESA-SY-0059, ESTEC, Noordwijk, The Netherlands, 81 pp.

Eymard L, Tabary L, Gerard E, Boukabara S-A, and Le Cornec A (1996) The microwave radiometer aboard ERS-1: Part II-Validation of the geophysical products. *IEEE Transactions on Geoscience and Remote Sensing* 34: 291–303.

Fleury R, Foucher F, and Lassudrie-Duchesne P (1991) Global TEC measurement capabilities of the DORIS system. *Advances in Space Research* 11: 51–54.

Fok HS, Baki Iz H, Shum CK, et al. (2010) Evaluation of ocean tide models used for Jason-2 altimetry corrections. *Marine Geodesy* 33(S1): 285–303.

Francis O and Bergé M (1993) Estimate of the radial orbit error by complex demodulation. *Journal of Geophysical Research* 98: 16083–16094.

Fu L-L (ed) (1998) Minutes of the joint TOPEX/POSEIDON Jason-1 science working team meeting. JPL Report D-16608, Jet Propul. Lab., Pasadena, CA, pp. 32–33.

Fu L-L, Alsdorf D, Morrow R, Rodriquez E, and Mognard N (2012) *SWOT: The Surface Water and Ocean Topography Mission, Wide-Swath Altimetric Measurement of Water Elevation on Earth*. Pasadena, CA: Jet Propulsion Laboratory/California Institute of Technology, JPL Publication 12–05, 228 pp.

Fu L-L and Cazanave A (eds.) (2001) *Satellite Altimetry and Earth Sciences*. In: *International Geophysics Series*, vol. 69. Academic Press.

Fu L-L, Christensen EJ, Lefebvre M, Ménard Y, Dorrer M, and Escudier P (1994) TOPEX/POSEIDON mission overview. *Journal of Geophysical Research* 99: 24369–24381.

Fu L-L and Glazman R (1991) The effect of the degree of wave development on the sea-state bias in radar altimetry measurement. *Journal of Geophysical Research* 96: 829–834.

Fukamori I, Raghunath R, and Fu L-L (1998) Nature of global large-scale sea level variability in relation to atmospheric forcing: A modeling study. *Journal of Geophysical Research* 103: 5493–5512.

Gaspar P and Florens JP (1998) Estimation of the sea state bias in radar altimeter measurements of sea level: Results from a new nonparametric method. *Journal of Geophysical Research* 103: 15803–15814.

Gaspar P, Labroue S, Ogor F, Lafitte G, Marchal L, and Rafanel M (2002) Improving nonparametric estimates of the sea state bias in radar altimeter measurements of sea level. *Journal of Atmospheric and Oceanic Technology* 19: 1690–1707.

Gaspar P, Ogor F, Le Traon P-Y, and Zanife OZ (1994) Estimating the sea state bias of the TOPEX and POSEIDON altimeters from crossover differences. *Journal of Geophysical Research* 99: 24981–24994.

Glazman R, Fabrikant A, and Srokosc M (1996) Numerical analysis of the sea state bias for satellite altimetry. *Journal of Geophysical Research* 101: 3789–3799.

Guman MD (1997) Determination of Global Mean Sea Level Variations Using Multi-Satellite Altimetry. PhD, Department of Aerospace Engineering, The University of Texas at Austin, 195 pp.

Gunter B, Urban T, Riva R, et al. (2009) A comparison of coincident GRACE and ICESat data over Antarctica. *Journal of Geodesy* 83: 1051–1060. http://dx.doi.org/10.1007/s00190-009-0323-4.

Haines BJ, Desai SD, and Born GH (2010) The harvest experiment: Calibration of the climate data record from TOPEX/Poseidon, Jason-1 and the Ocean Surface Topography Mission. *Marine Geodesy* 33(S1): 91–113.

Haines BJ, Dong D, Born GH, and Gill SK (2003) The Harvest experiment: Monitoring Jason-1 and TOPEX/Poseidon from a California Offshore Platform. *Marine Geodesy* 26: 239–260.

Hausman J and Zlotnicki V (2010) Sea state bias in radar altimetry revisited. *Marine Geodesy* 33(S1): 336–347.

Haxby WF, Karner GD, LaBrecque JL, and Weissel JK (1983) Digital images of combined oceanic and continental data sets and their use in tectonic studies. *Eos, Transactions American Geophysical Union* 64: 995–1004.

Hayne GS and Hancock DW (1982) Sea-state related altitude errors in the SEASAT radar altimeter. *Journal of Geophysical Research* 87: 3227–3231.

Hendricks JR, Leben RR, Born GH, and Koblinsky CJ (1996) Empirical orthogonal function analysis of global TOPEX/POSEIDON altimeter data and implications for detection of global sea level rise. *Journal of Geophysical Research* 101: 14131–14146.

Hevizi L, Walsh E, MacIntosh R, et al. (1993) Electromagnetic bias in sea surface range measurements at frequencies of the TOPEX/Poseidon satellite. *IEEE Transactions on Geoscience and Remote Sensing* 31: 367–388.

Holgate SJ (2007) On the decadal rates of sea level change during the twentieth century. *Geophysical Research Letters* 34: L01602. http://dx.doi.org/10.1029/2006GL028492.

Howat IM, Smith BE, Joughin I, and Scambos TA (2008) Rates of southeast Greenland ice volume loss from combined ICESat and ASTER observations. *Geophysical Research Letters* 35: L17505. http://dx.doi.org/10.1029/2008GL034496.

Imel DA (1994) Evaluation of the TOPEX/POSEDION dual-frequency ionosphere correction. *Journal of Geophysical Research* 99: 24895–24906.

Jacobs GA, Born GH, Parke ME, and Allen PC (1992) The global structure of the annual and semiannual sea surface height variability from the Geosat altimeter data. *Journal of Geophysical Research* 97: 17813–17828.

Jan G, Menard Y, Faillot M, Lyard F, Jeansou E, and Bonnefond P (2004) Offshore absolute calibration of space-borne radar altimeters. *Marine Geodesy* 27: 615–630.

Keihm SJ, Jansses MA, and Ruf CS (1995) TOPEX/POSEIDON microwave radiometer (TMR) III: Wet tropospheric range correction algorithm and pre-launch error budget. *IEEE Transactions on Geoscience and Remote Sensing* 33: 147–161.

Keihm SJ, Zlotnicki V, and Ruf CS (2000) TOPEX microwave radiometer performance evaluation. *IEEE Transactions on Geoscience and Remote Sensing* 38: 1379–1386.

Kichak RA (2003) Independent GLAS Anomaly Review Board Executive Summary, 4 November 2003. http://icesat.gsfc.nasa.gov/docs/IGARB.pdf.

Knudsen P, Bingham R, Andersen O, and Rio M-H (2011) A global mean dynamic topography and ocean circulation estimation using a preliminary GOCE gravity model. *Journal of Geodesy* 85: 861–879. http://dx.doi.org/10.1007/s00190-011-0485-8.

Kolenkiewicz R and Martin C (1982) Seasat altimeter height calibration. *Journal of Geophysical Research* 87: 3189–3197.

Labroue S, Boy F, Picot N, Urvoy M, and Ablain M (2012) First quality assessment of the Cryosat-2 altimetric system over ocean. *Advances in Space Research* 50: 1030–1045. http://dx.doi.org/10.1016/j.asr.2011.11.018.

Labroue S, Gaspar P, Dorandeu J, et al. (2006) Overview of the improvements made on the empirical determination of the sea state bias correction. In: *Proceedings of the 15 Years of Progress in Radar Altimetry Symposium, March 13–18, Venice, Italy*. http://earth.esa.int/workshops/venice06/participants/318/paper318labroue.pdf.

Labroue S, Gaspar P, Dorandeu J, et al. (2004) Nonparametric estimates of the sea state bias for the Jason-1 radar altimeter. *Marine Geodesy* 27: 453–483.

Lame DB and Born GH (1982) SEASAT measurement system evaluation: Achievements and limitations. *Journal of Geophysical Research* 87: 3175–3178.

Lemoine FG, Kenyon SC, Factor JK, et al. (1998) *The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96*. Greenbelt, MD: Goddard Space Flight Center, NASA TM-1998-206861.

Lemoine FG, Luthke SB, Zelensky NP, et al. (2003) An evaluation of recent gravity models with respect to altimeter satellite missions. In: *Presented at TOPEX/Poseidon and Jason-1 Science Working Team Meeting, Arles, France.*

Lerch FJ, Putney BH, Wagner CA, and Klosko SM (1981) Goddard earth models for oceanographic applications (GEM-10B and GEM-10C). *Marine Geodesy* 5: 145–187.

Leuliette EW, Nerem RS, and Mitchum GT (2004) Calibration of TOPEX/Poseidon and Jason altimeter data to construct a continuous record of mean sea level change. *Marine Geodesy* 27: 79–94.

Leuliette EW and Scharroo R (2010) Integrating Jason-2 into a multiple-altimeter climate data record. *Marine Geodesy* 33: 504–517.

Levitus S (1982) *Climatological Atlas of the World Ocean*. Rockville, MD: US Department of Commerce, NOAA Prof. Rep. 13.

Luthcke SB, Rowlands DD, Williams TA, and Sirota M (2005) Reduction of ICESat systematic geolocation errors and the impact on ice sheet elevation change detection. *Geophysical Research Letters* 32: L21S05. http://dx.doi.org/10.1029/2005GL023689.

Luthcke SB, Zelensky NP, Rowlands DD, Lemoine FG, and Williams TA (2003) The 1-centimeter orbit: Jason-1 precision orbit determination using GPS, SLR, DORIS, and altimeter data. *Marine Geodesy* 26: 399–421.

Ma XC, Shum CK, Eanes RJ, and Tapley BD (1994) Determination of ocean tides from the first year of TOPEX/POSEIDON altimeter measurements. *Journal of Geophysical Research* 99: 24809–24820.

Mannucci AJ, Iijima BA, Linqwister UJ, Pi X, Sparks L, and Wilson BD (1999) GPS and ionosphere. In: Stone WR (ed.) *Review of Radio Science*, pp. 622–655. Oxford: Oxford University Press.

Marsh JG, Koblinsky CJ, Zwally HJ, Brenner AC, and Beckley BD (1992) A global mean sea surface based upon GEOS-3 and Seasat altimeter data. *Journal of Geophysical Research* 97: 4915–4921.

Marsh JG and Martin TV (1982) The Seasat altimeter mean sea surface. *Journal of Geophysical Research* 87: 3269–3280.

Marsh JG and Williamson RG (1980) Precise orbit analysis in support of the Seasat altimeter experiment. *Journal of the Astronautical Sciences* 27: 345–369.

Martin C and Kolenkiewicz R (1981) Calibration validation of the GEOS-3 altimeter. *Journal of Geophysical Research* 86: 6369–6381.

Martinez-Benjamin JJ, et al. (2004) Ibiza absolute calibration experiment: Survey and preliminary results. *Marine Geodesy* 27: 657–682.

Mazzega P (1989) The solar tides and the sun-synchronism of satellite altimetry. *Geophysical Research Letters* 16: 507–510.

McConathy DR and Kilgus CC (1987) The Navy Geosat mission: And overview. *John Hopkins APL Technical Digest* 8(2): 170–175.

Melville WK, Stewart RH, Keller WC, et al. (1991) Measurements of electromagnetic bias in radar altimetry. *Journal of Geophysical Research* 96: 4915–4924.

Ménard Y, Jeansou E, and Vincent P (1994) Calibration of TOPEX/POSEIDON altimeters at Lampedusa: Additional results at Harvest. *Journal of Geophysical Research* 99: 24487–24504.

Merrifield MA (2011) A shift in western tropical Pacific sea level trends during the 1990s. *Journal of Climate* 24(15): 4126–4138.

Merrifield MA and Maltrud ME (2011) Regional sea level trends due to a Pacific trade wind intensification. *Geophysical Research Letters* 38: L21605. http://dx.doi.org/10.1029/2011GL049576.

Merrifield MA, Merrifield ST, and Mitchum GT (2009) An anomalous recent acceleration of global sea level rise. *Journal of Climate* 22: 5772–5781.

Merrifield MA, Thompson PR, and Lander M (2012) Multidecadal sea level anomalies and trends in the western tropical Pacific. *Geophysical Research Letters* 39: L13602. http://dx.doi.org/10.1029/2012GL052032.

Minster JF, Brossier C, and Rogel P (1995) Variation of the mean sea level from TOPEX/Poseidon data. *Journal of Geophysical Research* 100: 25153–25162.

Minster JF, Cazenave A, Serafini YV, Mercier F, Gennero MC, and Rogel P (1999) Annual cycle in mean sea level from TOPEX-Poseidon and ERS-1: Inference on the global hydrological cycle. *Global and Planetary Change* 20: 57–66.

Mitchum GT (1994) Comparison of TOPEX sea surface heights and tide gauge sea levels. *Journal of Geophysical Research* 99: 24541–24553.

Mitchum GT (1998) Monitoring the stability of satellite altimeters with tide gauges. *Journal of Atmospheric and Oceanic Technology* 15: 721–730.

Mitchum GT (2000) An improved calibration of satellite altimetric heights using tide gauge sea levels with adjustment for land motion. *Marine Geodesy* 23: 145–166.

Nerem RS (1995) Measuring global mean sea level variations using TOP EX/POSEIDON altimeter data. *Journal of Geophysical Research* 100: 25135–25151.

Nerem RS, Chambers DP, Choe C, and Mitchum GT (2010) Estimating mean sea level change from the TOPEX and Jason altimeter missions. *Marine Geodesy* 33: 435–446.

Nerem RS, Chambers DP, Leuliette E, Mitchum GT, and Giese BS (1999) Variations in global mean sea level during the 1997–98 ENSO event. *Geophysical Research Letters* 26: 3005–3008.

Nerem RS, Eanes RJ, Ries JC, and Mitchum GT (1998) *The use of a precise reference frame for sea level studies*. *Integrated Global Geodetic Observing System*. Munich: International Association of Geodesy.

Nerem RS, Haines BJ, Hendricks J, Minster JF, Mitchum GT, and White WB (1997) Improved determination of global mean sea level variations using TOPEX/POSEIDON altimeter data. *Geophysical Research Letters* 24: 1331–1334.

Nerem RS, Lerch FJ, Marshall JA, et al. (1994a) Gravity model development for TOPEX/POSEIDON: Joint gravity model-1 and 2. *Journal of Geophysical Research* 99: 24421–24447.

Nerem RS, Schrama EJ, Koblinsky CJ, and Beckley BD (1994b) A preliminary evaluation of ocean topography from the TOPEX/POSEIDON mission. *Journal of Geophysical Research* 99: 24565–24583.

Obligas E, Tran N, and Eymard L (2004) An assessment of Jason-1 microwave radiometer measurements and products. *Marine Geodesy* 27: 255–277.

Obligis E, Eymard L, and Tran N (2007) A new sidelobe correction algorithm for microwave radiometers: Application to the Envisat instrument. *IEEE Transactions on Geoscience and Remote Sensing* 43: 602–612.

Parke ME, Born G, Leben R, McLaughlin C, and Tierney C (1998) Altimeter sampling characteristics using a single satellite. *Journal of Geophysical Research* 103: 10513–10526.

Parke ME, Stewart RH, Farless DL, and Cartwright DE (1987) On the choice of orbits for an altimetric satellite to study ocean circulation and tides. *Journal of Geophysical Research* 92: 11693–11707.

Pavlis NK, Holmes SA, Kenyon SC, and Factor JK (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research* 117: B04406. http://dx.doi.org/10.1029/2011JB008916.

Pavlis EC and Mertikas SP (2004) The GAVDOS mean sea level and altimeter calibration facility: Results for Jason-1. *Marine Geodesy* 27: 631–656.

Ponte RM, Salstein DA, and Rosen RD (1991) Sea level response to pressure forcing in a barotropic numerical model. *Journal of Physical Oceanography* 21: 1043–1057.

Pritchard HD, Arthern RJ, Vaughan DG, and Edwards LA (2009) Extensive dynamic thinning of the margins of the Greenland and Antarctic ice sheets. *Nature* 461: 971–975. http://dx.doi.org/10.1038/nature08471.

Raney RK (1998) The delay Doppler radar altimeter. *IEEE Transactions on Geoscience and Remote Sensing* 36: 1578–1588.

Rapp RH (1986) Gravity anomalies and sea surface heights derived from a combined GEOS-3/Seasat altimeter data set. *Journal of Geophysical Research* 91: 4867–4876.

Rapp RH (1989) The treatment of permanent tidal effects in the analysis of satellite altimeter data for sea surface topography. *Manuscripta Geodaetica* 14: 368–372.

Ray RD and Douglas BC (2011) Experiments in reconstructing twentieth-century sea levels. *Progress in Oceanography* 91(4): 496–515.

Rodriguez E, Kim Y, and Martin YM (1992) The effect of small-wave modulation on the electromagnetic bias. *Journal of Geophysical Research* 97: 2379–2389.

Rodriguez E and Martin YM (1992) Theory and design of interferometric synthetic aperture radars. *IEE Proceedings F, Radar and Signal Processing* 139: 147–159.

Rodriguez E and Martin YM (1994) Estimation of the electromagnetic bias from retracked TOPEX data. *Journal of Geophysical Research* 99: 24971–24979.

Rodriguez E and Pollard BD (2001) The measurement capabilities of wide-swath ocean altimeters. In: Chelton DB (ed.) *Report of the High-Resolution Ocean Topography Science Working Group Meeting*, pp. 190–215. Corvallis, OR: College of Oceanic and Atmospheric Sciences, Oregon State University.

Ruf CS (2002) Characterization and correction of a drift in calibration of the TOPEX microwave radiometer. *IEEE Transactions on Geoscience and Remote Sensing* 40: 509–511.

Ruf C, Keihm S, Subramanya B, and Jansses M (1994) TOPEX/POSEIDON microwave radiometer performance and in-flight calibration. *Journal of Geophysical Research* 99: 24915–24926.

Rummel R and Gruber T (2010) Gravity and steady-state ocean circulation explorer. In: Flechtner F, et al. (ed.) *System Earth via Geodetic-Geophysical Space Techniques*, pp. 203–212. Berlin, Heidelberg: Springer.

Sandwell DT (1992) Antarctic marine gravity field from high-density satellite altimetry. *Reviews of Geophysical Supplement* 29: 132–137.

Sandwell DT and Smith WHF (1997) Marine gravity anomaly from Geosat and ERS-1 satellite altimetry. *Journal of Geophysical Research* 102: 10039–10054.

Sandwell DT and Smith WHF (2001) Bathymetric estimation. In: Fu L-L and Cazanave A (eds.) *Satellite Altimetry and Earth Science. International Geophysics Series*, vol. 69, pp. 441–457. Academic Press, Chapter 12.

Scharroo R (2002) *A Decade of ERS Satellite Orbits and Altimetry*. PhD Dissertation, The Netherlands: Department of Aerospace Engineering, Delft University of Technology, 195 pp.

Scharroo R, Lillibridge JL, Smith WHF, and Schrama EJO (2004) Cross-calibration and long-term monitoring of the microwave radiometers of ERS, TOPEX, GFO, Jason, and Envisat. *Marine Geodesy* 27: 279–297.

Scharroo R and Visser PNAM (1998) Precise orbit determination and gravity field improvement for the ERS satellites. *Journal of Geophysical Research* 103: 8113–8127.

Schlax MG and Chelton DB (1994) Aliased tidal errors in TOPEX/POSEIDON sea surface height data. *Journal of Geophysical Research* 99: 24725–24761.

Schrama EJO and Ray RD (1994) A preliminary tidal analysis of TOPEX/POSEIDON altimetry. *Journal of Geophysical Research* 99: 24799–24808.

Schreiner WS, Markin RE, and Born GH (1997) Correction of signal frequency altimeter measurements for ionospheric delay. *IEEE Transactions on Geoscience and Remote Sensing* 35: 271–277.

Schutz BE and Tapley BD (1980) Orbit accuracy assessment for Seasat. *Journal of the Astronautical Sciences* 27: 371–390.

Schutz BE, Zwally HJ, Shuman CA, Hancock D, and DiMarzio JP (2005) Overview of the ICESat mission. *Geophysical Research Letters* 32: L21S01. http://dx.doi.org/10.1029/2005GL024009.

Shepherd A, Wingham D, Ridout A, and Muir A (2010) Initial assessment of CryoSat-2 performance over continental ice (invited). In: *Presented at Fall Meeting of American Geophysical Union, San Francisco, CA*.

Shum CK, Chambers DP, Ries JC, Tapley BD, and Yuan DN (1994) *Determination of large scale ocean circulation and its variations using satellite altimetry. Gravimetry and Space Techniques Applied to Geodynamics and Ocean Dynamics. Geophysical Monographic Series*, vol. 82. Washington, DC: AGU, 21–32.

Shum CK, Tapley BD, and Ries JC (1993) Satellite altimetry: Its applications and accuracy assessment. *Advances in Space Research* 13(11): 315–324.

Shum C, Yi Y, Cheng K, Kuo C, Braun A, Calmant S, and Chambers D (2003) Calibration of Jason-1 altimeter over Lake Erie. *Marine Geodesy* 26: 335–354.

Sibthorpe A, Brown S, Desai SD, and Haines BJ (2011) Calibration and validation of the Jason-2/OSTM advanced microwave radiometer using terrestrial GPS stations. *Marine Geodesy* 34: 420–430.

Smith TM, Reynolds RW, Livezey RE, and Stokes DC (1996) Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *Journal of Climate* 9: 1403–1420.

Smith WHF and Sandwell DT (1994) Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry. *Journal of Geophysical Research* 99: 21803–21824.

Smith DE, Zuber MT, Frey HV, et al. (2001) Mars orbiter laser altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research* 106: 23689–23722.

Sorensen LS, Simonsen SB, Nielsen K, et al. (2011) Mass balance of the Greenland ice sheet (2003–2008) from ICESat data – the impact of interpolation, sampling and firn density. *The Cryosphere* 5: 173–186. www.the-cryosphere.net/5/173/2011/ http://dx.doi.org/10.5194/tc-5-173-201.

Srokosz MA (1986) On the joint distribution of surface elevations and slopes for a nonlinear random sea, with an application for radar altimetry. *Journal of Geophysical Research* 91: 995–1006.

Stammer D, Wunsch C, and Ponte RM (2000) De-aliasing of global high frequency barotropic motions in altimeter measurements. *Geophysical Research Letters* 27: 1175–1178.

Stanley HR (1979) The GEOS-3 project. *Journal of Geophysical Research* 84(B8): 3779–3783.

Stewart RH, Fu L-L, and Lefebvre M (1986) *Science opportunities from the TOPEX/Poseidon mission*. Pasadena, CA: Jet Propulsion Laboratory Publication, 86–18, 58 pp.

Tai CK (1988) Geosat crossover analysis in the tropical pacific 1. Constrained sinusoidal crossover adjustment. *Journal of Geophysical Research* 93: 10621–10629.

Tapley BD, Bettadpur S, Watkins M, and Riegber C (2004a) The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters* 31: L09607, 1–4.

Tapley BD, Born GH, and Parke ME (1982) The SEASAT altimeter data and its accuracy assessment. *Journal of Geophysical Research* 87: 3179–3188.

Tapley BD, Chambers DP, Bettadpur S, and Ries JC (2003) Large scale ocean circulation from the GRACE GGM01 Geoid. *Geophysical Research Letters* 30. http://dx.doi.org/10.1029/2003GL018622.

Tapley BD, Chambers DP, Shum CK, Eanes RJ, Ries JC, and Stewart RH (1994a) Accuracy assessment of the large-scale dynamic ocean topography from TOPEX/POSEIDON altimetry. *Journal of Geophysical Research* 99: 24605–24618.

Tapley BD and Kim MC (2001) Applications to geodesy. In: Fu L-L and Cazanave A (eds.) *Satellite Altimetry and Earth Science. International Geophysics Series*, vol. 69, pp. 371–406. Academic Press, Chapter 10.

Tapley B, Ries J, Bettadpur S, et al. (2005) GGM02 – An improved Earth gravity field model from GRACE. *Journal of Geodesy* http://dx.doi.org/10.1007/s00190-005-0480-z.

Tapley BD, Ries JC, Davis GW, et al. (1994b) Precision orbit determination for TOPEX/POSEIDON. *Journal of Geophysical Research* 99(C12): 24383–24404.

Tapley BD and Rosborough GW (1985) Geographically correlated orbit errors and its effect on satellite altimetry missions. *Journal of Geophysical Research* 90: 11817–11831.

Tapley BD, Schutz BE, and Born GH (2004b) *Statistical Orbit Determination*. Academic Press.

Tapley BD, Shum CK, Ries JC, Suter R, and Schutz BE (1992) Monitoring of changes in global mean sea level using Geosat altimeter. *Sea Level Changes: Determination and Effects. Geophysical Monograph Series*, vol. 69, pp. 167–180. Washington: IUGG.

Tapley BD, Shum CK, Ries JC, et al. (1997) The TEG-3 geopotential model. In: Seagwa J, Fujimoto H, and Okubo S (eds.) *Gravity, Geoid and Marine Geodesy*. Proceedings of the International Association of Geodesy International Symposium No. 117, pp. 453–460. New York: Springer.

Tapley BD, Watkins MM, Ries JC, et al. (1996) The JGM-3 geopotential model. *Journal of Geophysical Research* 101(B12): 28029–28049.

Tierney C, Wahr J, Bryan F, and Zlotniki V (2000) Short-period oceanic circulation: Implications for satellite altimetry. *Geophysical Research Letters* 27: 1255–1258.

Tran N, Labroue S, Philipps S, Bronner E, and Picot N (2010) Overview and update of the sea state bias corrections for the Jason-2, Jason-1 and TOPEX missions. *Marine Geodesy* 33(S1): 348–362.

Tran N, Vandemark D, Chapron B, et al. (2006) New models for satellite altimeter sea state bias correction developed using global wave model data. *Journal of Geophysical Research* 111: c09009. http://dx.doi.org/10.1029/2005JC003406.

Trenberth KE and Olson JG (1988) An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Center for Medium-Range Weather Forecasts. *Bulletin of the American Meteorological Society* 69: 1047–1057.

Urban T (2000) The Integration and Application of Multi-Satellite Radar Altimetry. PhD, Department of Aerospace Engineering, The University of Texas at Austin, 273 pp.

Urban TJ and Schutz BE (2005) ICESat sea level comparisons. *Geophysical Research Letters* 32: L23S10. http://dx.doi.org/10.1029/2005GL024306.

van Gysen H, Coleman R, Morrow R, Hirsch B, and Rizos C (1992) Analysis of collinear passes of satellite altimeter data. *Journal of Geophysical Research* 97: 2265–2277.

Vandemark D, Tran N, Beckley BD, Chapron B, and Gaspar P (2002) Direct estimation of sea state impacts on radar altimeter sea level measurements. *Geophysical Research Letters* 29: 2148. http://dx.doi.org/10.1029/2002GL015776.

Wagner CA and Cheney RE (1992) Global sea level changes from satellite altimetry. *Journal of Geophysical Research* 97: 15607–15615.

Wagner CA and Tai CK (1994) Degradation of ocean signals in satellite altimetry due to orbit error removal processes. *Journal of Geophysical Research* 99: 16255–16267.

Walsh EJ, Jackson FC, Hines DE, et al. (1991) Observations on electromagnetic bias in radar altimeter surface range measurements. *Journal of Geophysical Research* 96: 20571–20583.

Wang YM (2000) The satellite altimeter data derived mean sea surface GSFC98. *Geophysical Research Letters* 27: 701–704.

Watson C, Coleman R, White N, Church J, and Govind R (2003) Absolute calibration of TOPEX/Poseidon and Jason-1 using GPS buoys in Bass Strait, Australia. *Marine Geodesy* 26: 285–304.

Willis JK, Roemmich D, and Cornuelle B (2004) Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales. *Journal of Geophysical Research* 109: C12036. http://dx.doi.org/10.1029/2003JC002260.

Witter DL and Chelton DB (1991) An apparent wave height dependence in the sea-state bias in GEOSAT altimeter range measurements. *Journal of Geophysical Research* 96: 8861–8867.

Woodworth P, Moore P, Dong X, and Bingley R (2004) Absolute calibration of the Jason-1 altimeter using UK tide gauges. *Marine Geodesy* 27: 95–106.

Wunsch C (1972) Bermuda sea level in relation to tides, weather and baroclinic fluctuations. *Reviews of Geophysical Space Research* 10: 1–49.

Wunsch C and Gaposchkin EM (1980) On using satellite altimetry to determine the general circulation of the oceans with application to geoid improvement. *Reviews of Geophysics* 18: 725–745.

Wyrtki K and Mitchum G (1990) Interannual differences of Geosat altimeter heights and sea level: The importance of a datum. *Journal of Geophysical Research* 95: 2969–2975.

Zaouche G, Perbos J, Lafon T, et al. (2010) OSTM/Jason-2: Assessment of the system performances (Ocean Surface Topography Mission: OSTM). *Marine Geodesy* 33: 26–52.

Zimbelman DF and Busalacchi AJ (1990) The wet tropospheric range correction: Product intercomparisons and the simulated effect for tropical Pacific altimeter retrievals. *Journal of Geophysical Research* 95: 2899–2922.

Zlotnicki V (1994) Correlated environmental corrections for TOPEX/POSEIDON, with a note of ionospheric accuracy. *Journal of Geophysical Research* 99: 24907–24914.

Zlotnicki V and Desai SD (2004) Assessment of Jason microwave radiometer's measurement of wet tropospheric path delay using comparisons with SSM/I and TMI. *Marine Geodesy* 27: 241–253.

Zwally HJ and Brenner AC (2001) Ice sheet dynamics and mass balance. In: Fu L-L and Cazanave A (eds.) *Satellite Altimetry and Earth Science. International Geophysics Series*, vol. 69, pp. 351–369. Academic Press, Chapter 9.

Zwally HJ, Schutz B, Abdalati W, et al. (2002) ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics* 34: 405–445.

This page intentionally left blank

3.06 Earth Tides

DC Agnew, University of California, La Jolla, CA, USA

© 2015 Elsevier B.V. All rights reserved.

3.06.1	**Introduction**	151
3.06.1.1	An Overview	152
3.06.2	**The Tidal Forces**	152
3.06.2.1	The Tidal Potential (I)	153
3.06.2.2	Computing the Tides: Ephemeris-Based	154
3.06.2.3	Computing the Tides (II): Harmonic Decompositions	155
3.06.2.4	The Pole Tide	157
3.06.2.5	Radiational Tides	159
3.06.3	**Tidal Response of the Solid Earth**	159
3.06.3.1	Tidal Response of an SNREI Earth	159
3.06.3.1.1	Some combinations of Love numbers (I): gravity and tilt	160
3.06.3.1.2	Combinations of Love numbers (II): Displacement and strain tides	160
3.06.3.2	Response of a Rotating Earth	161
3.06.3.2.1	NDFW resonance	161
3.06.3.2.2	Coupling to other modes	163
3.06.3.2.3	Anelastic effects	164
3.06.3.3	Lateral Heterogeneities	165
3.06.4	**Tidal Loading**	165
3.06.4.1	Computing Loads I: Spherical Harmonic Sums	165
3.06.4.2	Computing Loads II: Integration Using Green Functions	166
3.06.4.3	Degree-One Effects in Ocean Loading	168
3.06.4.4	Ocean Tide Models	169
3.06.4.5	Computational Methods	169
3.06.5	**Analyzing and Predicting Earth Tides**	169
3.06.5.1	Tidal Analysis and Prediction	169
3.06.5.1.1	Predicting tides	171
3.06.6	**Earth-Tide Instruments and Measurements**	172
3.06.6.1	Local Distortion of the Tides	173
References		174

3.06.1 Introduction

The Earth tides are the motions in the solid Earth, and the changes in its gravitational potential, induced by the tidal forces from external bodies (these forces, acting on the rotating Earth, also induce motions of its spin axis; see **Chapter 3.09**). Earth tides have three roles in geophysics: measurements of them can provide information about the Earth; models of them can be used to remove tidal variations from measurements; and the same models can be used to examine tidal influence on some phenomenon. An example of the first role would be measuring the nearly diurnal resonance in the gravity tide to estimate the flattening of the core–mantle boundary (**Section 3.06.3.2.1**); of the second, computing the expected tidal displacements at a point so we can better estimate its position with GPS; and of the third, finding the tidal stresses to see if they trigger earthquakes. For the last two activities, we need to be able to model the Earth tides very accurately, something that is in fact not difficult. Notably, Earth tides are much easier to model accurately than ocean tides, both because the Earth is more rigid than water and because the geometry of the problem is much simpler.

Accurate modeling of Earth tides is also easy because they do not depend very strongly on Earth structure – which also means that Earth-tide measurements, and accurate estimates of the few parameters that describe the Earth's response to the tides, do not provide much information about the Earth. This was not always true; for example, in 1922, Jeffreys used tidal data to show that the average rigidity of the Earth was much less than that of the mantle, indicating that the core must be of very low rigidity (Brush, 1996). But since then, seismology has determined Earth structure in much more detail than could be found using tides. Recently, Earth tides have become more important in geodesy, as the increasing precision of measurements has required corrections for previously undetectable tidal effects. This chapter focuses on the theory needed to compute tidal effects as accurately as possible for understanding how phenomena are or are not influenced by tides. **Section 3.06.6** summarizes, more briefly, measurement techniques and some results.

The best short introduction to Earth tides remains that of Baker (1984). Melchior (1983) described the subject fully (and with a very complete bibliography) but is now somewhat out of date and should be used with caution by the newcomer to

the field. The volume of articles edited by Wilhelm et al. (1997) is a better reflection of the current state of the subject, as are the quadrennial International Symposia on Earth Tides (e.g., Jentzsch et al., 2009). Harrison (1985) reprinted a number of important papers, with very thoughtful commentary; Cartwright (1999) is a history of tidal science (mostly the ocean tides) that also provides an interesting introduction to some aspects of the field – which, as one of the older parts of geophysics, has a terminology sometimes overly affected by its history.

3.06.1.1 An Overview

Figure 1 is a simple flowchart to indicate what goes into a tidal signal. We usually take the tidal forcing to be completely known, but it is computed using a particular theory of gravity, and Earth-tide measurements actually provided some of the best evidence available for general relativity as opposed to some alternative theories (Warburton and Goodkind, 1976). The large box labeled 'Geophysics/Oceanography' includes the response of the Earth and ocean to the forcing, with the arrow going around it to show that some tidal signals would be observed even if the Earth were oceanless and rigid. Finally, measurements of Earth tides can detect other environmental and tectonic signals.

At this point, it is useful to introduce some terminology. The *theoretical tides* could be called the modeled tides, since they are computed from a set of models. The first model is the tidal forcing, or *equilibrium tidal potential*, produced by external bodies; this is computed from gravitational and astronomical theory and is the tide at point E in Figure 1. The next two models are those that describe how the Earth and ocean respond to this forcing; in Figure 1, these are boxes inside the large dashed box. The solid-Earth model gives what are called the *body tides*, which are what would be observed on an oceanless but otherwise realistic Earth. The ocean model (which includes both the oceans and the elastic Earth) gives the *load tides*, which are changes in the solid Earth caused by the shifting mass of water associated with the ocean tides. These two responses sum to give the total tide caused by the nonrigidity of the Earth; the final model, labeled 'Site distortions,' may be used to describe how local departures from idealized models affect the result (Section 3.06.6.1). This nonrigid contribution is summed with the tide from direct attraction to give the total theoretical tide, at point T in the flowchart.

Mathematically, we can describe the processes shown on this flow chart in terms of linear systems, something first applied to tidal theory by Munk and Cartwright (1966). The total signal $y(t)$ is represented by a convolution integral:

$$y(t) = \int x_T(t-\tau) w_T(\tau) d\tau + n(t) \qquad [1]$$

where $x_T(t)$ is the tidal forcing, $n(t)$ is the noise (nontidal energy, from whatever source), and τ is a dummy variable. The function $w(t)$ is the impulse response of the system to the tidal forcing. Fourier transforming equation [1], and disregarding the noise, gives $Y(f) = W_T(f) X_T(f)$: $W(f)$ is the *tidal admittance*, which turns out to be more useful than $w(t)$, partly because of the bandlimited nature of X_T and also because, with the exception of the NDFW resonance discussed in Section 3.06.3.2.1, $W(f)$ is a fairly smooth function of frequency. To predict the tides, we assume $W(f)$ (perhaps guided by previous measurements); to analyze them, we determine $W(f)$.

We describe the tidal forcing first, in some detail because the nature of this forcing governs the response and how tidal measurements are analyzed. We next consider how the solid Earth responds to the tidal forcing and what effects this produces. After this, we discuss the load tides, completing what we need to know to produce the full theoretical tides. We conclude with brief descriptions of analysis methods appropriate to Earth-tide data and instruments for measuring Earth tides.

3.06.2 The Tidal Forces

The tidal forces arise from the gravitational attraction of bodies external to the Earth. As noted earlier, computing them requires only some gravitational potential theory and astronomy,

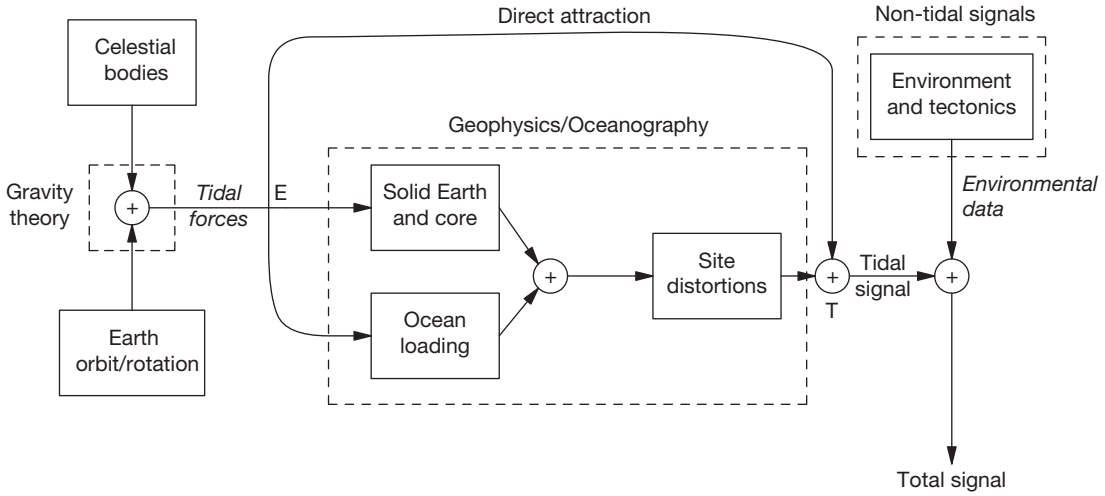


Figure 1 Tidal flowchart. Entries in italics represent things we know (or think we know) to very high accuracy; entries in boldface (over the dashed boxes) represent things we can learn about using tidal data. See text for details.

with almost no geophysics. The extraordinarily high accuracy of celestial mechanics makes it easy to describe the tidal forcing to much higher precision than can be measured: in this part of the subject, the romance of the next decimal place has exerted a somewhat excessive pull.

Our formal derivation of the tidal forcing will use potential theory, but it is useful to start by considering the gravitational forces exerted on one body (the Earth, in this case) by another. As usual in discussing gravitation, we work in terms of accelerations. Put most simply, the tidal acceleration at a point on or in the Earth is the difference between the acceleration caused by the attraction of the external body and the orbital acceleration – which is to say, the acceleration that the Earth undergoes as a whole. This result is valid whatever the nature of the orbit – it would hold just as well if the Earth were falling directly toward another body. For a spherically symmetrical Earth, the orbital acceleration is the acceleration caused by the attraction of the other body at the Earth's center of mass, making the tidal force the difference between the attraction at the center of mass and that at the point of observation.

Figure 2 shows the resulting force field. At the point directly under the attracting body (the *subbody point*), and at its antipode, the tidal force is oppositely directed in space, though in the same way (up) viewed from the Earth. It is in fact larger at the subbody point than at its antipode, though if the ratio a/R is small (1/60 for the forces plotted in Figure 2), the difference is also small.

3.06.2.1 The Tidal Potential (I)

We now derive an expression for the tidal force – or rather, for the more useful *tidal potential*, following the development in Munk and Cartwright (1966). If M_{ext} is the mass of the external body, the gravitational potential, V_{tot}, from it at O is

$$V_{tot} = \frac{GM_{ext}}{\rho} = \frac{GM_{ext}}{R}\frac{1}{\sqrt{1+(a/R)^2-2(a/R)\cos\alpha}}$$

using the cosine rule from trigonometry. The variables are as shown in Figure 2: r is the distance of O from C, ρ the distance from O to M, and α the angular distance between O and the subbody point of M. We can write the square-root term as a sum of Legendre polynomials, using the generating-function expression for these, which yields

$$V_{tot} = \frac{GM_{ext}}{R}\sum_{n=0}^{\infty}\left(\frac{a}{R}\right)^n P_n(\cos\ \alpha) \qquad [2]$$

where $P_2(x) = \frac{1}{2}(3x^2-1)$ and $P_3(x) = \frac{1}{2}(5x^3-3x)$.

The $n=0$ term is constant in space, so its gradient (the force) is zero, and it can be discarded. The $n=1$ term is

$$\frac{GM_{ext}}{R^2}a\cos\alpha = \frac{GM_{ext}}{R^2}x_1 \qquad [3]$$

where x_1 is the Cartesian coordinate along the C–M axis. The gradient of this is a constant, corresponding to a constant force along the direction to M; but this is just the orbital force at C, which we subtract to get the tidal force. Thus, the tidal potential is eqn [2] with the two lowest terms removed:

$$V_{tid}(t) = \frac{GM_{ext}}{R(t)}\sum_{n=2}^{\infty}\left(\frac{a}{R(t)}\right)^n P_n[\cos\alpha(t)] \qquad [4]$$

where we have made R and α, as they actually are, functions of time t – which makes V such a function as well.

We can now put in some numbers appropriate to the Earth and relevant external bodies to get a sense of the magnitudes of different terms. If r is the radius of the Earth, $a/R = 1/60$ for the Moon, so that the size of terms in the sum [4] decreases fairly rapidly with increasing n; in practice, we need only consider $n=2$ and $n=3$, and perhaps $n=4$ for the highest precision; the $n=4$ tides are just detectable in very low-noise gravimeters. These different values of n are referred to as the *degree-n* tides. For the Sun, $r/R = 1/23\,000$, so the degree-2 solar tides completely dominate.

If we consider $n=2$, the magnitude of V_{tid} is proportional to GM_{ext}/R^3. If we normalize this quantity to make the value for the Moon equal to 1, the value for the Sun is 0.46, for Venus 5×10^{-5}, for Jupiter 6×10^{-6}, and even less for all other planets. So the *lunisolar* tides dominate and are probably the only ones visible in actual measurements – though, as we will see, some expansions of the tidal potential include planetary tides.

At very high precision, we also need to consider another small effect: the acceleration of the Earth is exactly equal to the attraction of the external body at the center of mass only for a spherically symmetrical Earth. For the real Earth, the C_{20} term in the gravitational potential makes the acceleration of the Moon by the Earth (and hence the acceleration of the Earth by the Moon) depend on more than just eqn [3]. The resulting

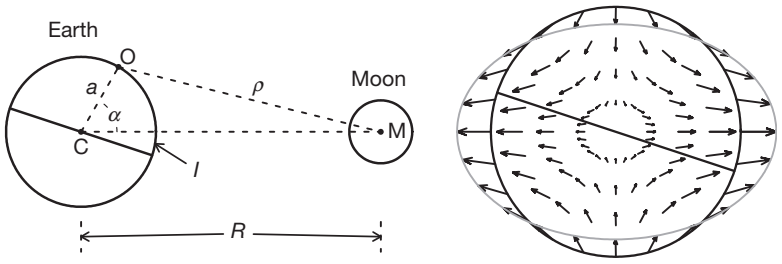


Figure 2 Tidal forcing. On the left is the geometry of the problem for computing the tidal force at a point O on the Earth, given an external body M. The right plot shows the field of forces (accelerations) for the actual Earth–Moon separation; the scale of the largest arrow is 1.14 μm s^{-2} for the Moon, and 0.51 μm s^{-2} for the Sun. The gray line shows the equipotential surface under tidal forcing, greatly exaggerated. The solid line is the Earth's equator; I is the inclination angle of the Moon.

Earth-flattening tides (Dahlen, 1993; Wilhelm, 1983) are very small.

We obtain further insight on the behavior of the tidal forces if we use geographic coordinates, rather than angular distance from the subbody point. Suppose our observation point O is at colatitude θ and east longitude ϕ (which are fixed) and that the subbody point of M is at colatitude $\theta'(t)$ and east longitude $\phi'(t)$. Then, we may apply the addition theorem for spherical harmonics to get, instead of [4],

$$V_{\text{tid}} = \frac{GM_{\text{ext}}}{R(t)} \sum_{n=2}^{\infty} \left(\frac{a}{R(t)} \right)^n \frac{4\pi}{2n+1} \sum_{m=-n}^{n} Y_{nm}^*(\theta'(t), \phi'(t)) Y_{nm}(\theta, \phi) \quad [5]$$

where we have used the fully normalized complex spherical harmonics defined by

$$Y_{nm}(\theta, \phi) = N_n^m P_n^m(\cos\theta) e^{im\phi}$$

where N_n^m is the normalizing factor

$$N_n^m = (-1)^m \left[\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} \right]^{\frac{1}{2}}$$

and P_n^m is the associated Legendre polynomial of degree n and order m (**Table 1**).

As is conventional, we express the tidal potential as V_{tid}/g, where g is the Earth's gravitational acceleration; this combination has the dimension of length and can easily be interpreted as the change in elevation of the geoid or of an equilibrium surface such as an ideal ocean (hence its name, the *equilibrium potential*.) Part of the convention is to take g to have its value on the Earth's equatorial radius a_{eq}; if we hold r fixed at that radius in [5], we get

$$\begin{aligned} \frac{V_{\text{tid}}}{g} &= a_{\text{eq}} \frac{M_{\text{ext}}}{M_{\text{E}}} \sum_{n=2}^{\infty} \frac{4\pi}{2n+1} \left(\frac{a_{\text{eq}}}{R} \right)^{n+1} \sum_{m=-n}^{n} Y_{nm}^*(\theta', \phi') Y_{nm}(\theta, \phi) \\ &= \sum_{n=2}^{\infty} K_n \frac{4\pi}{2n+1} \xi^{n+1} \sum_{m=-n}^{n} Y_{nm}^*(\theta', \phi') Y_{nm}(\theta, \phi) \end{aligned} \quad [6]$$

where the constant K includes all the physical quantities:

$$K_n = a_{\text{eq}} \frac{M_{\text{ext}}}{M_{\text{E}}} \left(\frac{a_{\text{eq}}}{\overline{R}} \right)^{n+1}$$

where M_{E} is the mass of the Earth and $\overline{R}$ is the mean distance of the body; the quantity $\xi = \overline{R}/R$ expresses the normalized variation in distance. For the Moon, K_2 is 0.35837 m, and for the Sun, 0.16458 m.

In both [5] and [6], we have been thinking of θ and ϕ as giving the location of a particular place of observation; but if we consider them to be variables, the $Y_{nm}(\theta, \phi)$ describes the geographic distribution of V/g on the Earth. The time dependence of the tidal potential comes from time variations in ξ, θ', and ϕ'. The first two change relatively slowly because of the orbital motion of M around the Earth; ϕ' varies approximately daily as the Earth rotates beneath M. The individual terms in the sum over m in [6] thus separate the tidal potential of degree n into parts, called *tidal species*, that vary with frequencies around 0, 1, 2, ... n times per day; for the largest tides ($n = 2$), there are three such species. The diurnal tidal potential varies once per day and with colatitude as $\sin\theta\cos\theta$: it is largest at midlatitudes and vanishes at the equator and the poles. The semidiurnal part (twice per day) varies as $\sin^2\theta$ and so is largest at the equator and vanishes at the poles. The long-period tide varies as $3\cos^2\theta - 1$ and so is large at the pole and (with reversed sign) at the equator. As we will see, these spatial dependences do not carry over to those tides, such as strain and tilt, that depend on horizontal gradients of the potential.

Table 1 Associated Legendre functions

$P_2^0(\theta) = \frac{1}{2}\left(3\cos^2\theta - 1\right)$	$P_3^0(\theta) = \frac{1}{2}\left(5\cos^3\theta - 3\cos\theta\right)$
$P_2^1(\theta) = 3\ \sin\theta\cos\theta$	$P_3^1(\theta) = \frac{3}{2}\left(5\cos^2\theta - 1\right)$
$P_2^2(\theta) = 3\ \sin^2\theta$	$P_3^2(\theta) = 15\ \sin^2\theta\ \cos\theta$
	$P_3^3(\theta) = 15\ \sin^3\theta$

To proceed further, it is useful to separate the time-dependent and space-dependent parts a bit more explicitly. We adopt the approach of Cartwright and Tayler (1971) who produced what was for a long time the standard harmonic expansion of the tidal potential. We can write [6] as

$$\begin{aligned} \frac{V_{\text{tid}}}{g} &= \sum_{n=2}^{\infty} K_n \xi^{n+1} \frac{4\pi}{2n+1} [Y_{n0}(\theta', \phi') Y_{n0}(\theta, \phi) \\ &\quad + \sum_{m=1}^{n} Y_{n-m}^*(\theta', \phi') Y_{n-m}(\theta, \phi) + Y_{nm}^*(\theta', \phi') Y_{nm}(\theta, \phi)] \\ &= \sum_{n=2}^{\infty} K_n \xi^{n+1} \frac{4\pi}{2n+1} [Y_{n0}(\theta', \phi') Y_{n0}(\theta, \phi) \\ &\quad + \sum_{m=1}^{n} 2\Re[Y_{nm}^*(\theta', \phi') Y_{nm}(\theta, \phi)]] \end{aligned}$$

Now, define complex (and time-varying) coefficients $T_{nm}(t) = a_n^m(t) + ib_n^m(t)$ such that

$$\frac{V_{\text{tid}}}{g} = \Re\left[\sum_{n=2}^{\infty} \sum_{m=0}^{n} T_{nm}^*(t) Y_{nm}(\theta, \phi) \right] \quad [7]$$

$$= \sum_{n=2}^{n=\infty} \sum_{m=0}^{n} N_n^m P_n^m(\cos\theta) \left[a_n^m(t) \cos m\phi + b_n^m(t) \sin m\phi \right] \quad [8]$$

Then, the T_{nm} coefficients are, for m equal to 0,

$$T_{n0} = \left(\frac{4\pi}{2n+1} \right)^{\frac{1}{2}} K_n \xi^{n+1} P_n^0(\cos\theta') \quad [9]$$

and, for m not equal to 0,

$$T_{nm} = (-1)^m \frac{8\pi}{2n+1} K_n \xi^{n+1} N_n^m P_n^m(\theta') e^{i\phi'} \quad [10]$$

from which we can find the real-valued, time-varying quantities $a_n^m(t)$ and $b_n^m(t)$, which we will use later in computing the response of the Earth.

3.06.2.2 Computing the Tides: Ephemeris-Based

Equations [7]–[9] suggest a straightforward way to compute the tidal potential (and, as we will see, other theoretical tides). First, use a description of the location of the Moon and Sun in

celestial coordinates (an ephemeris); other planets can be included if we wish. Then, convert this celestial location to the geographic coordinates θ' and ϕ' of the subbody point and the distance R, using standard transformations (Pétit and Luzum, 2010, **Chapter 1.05**). Finally, use eqns [9] and [10] to get $T_{nm}(t)$. Once we have the T_{nm}, we can combine these with the spatial factors in eqn [7] to get V_{tid}/g, either for a specific location or as a distribution over the whole Earth. Also, having done the astronomy to produce the T_{nm}, we can vary the spatial factors to find, not just the potential, but other observables, including tilt and strain.

This direct computation has the advantage, compared with the harmonic methods (discussed in the succeeding text), of being limited in accuracy only by the quality of the ephemeris. If we take derivatives of eqn [10] with respect to R, θ' and ϕ', we find that relative errors of 10^{-4} in V_{tid}/g would be caused by errors of 7×10^{-5} rad (14″) in θ' and ϕ' and 3×10^{-5} in ξ. (We pick this level of error because it usually exceeds the accuracy with which the tides can be measured, either because of noise or because of instrument calibration.) The errors in the angular quantities correspond to errors of about 400 m in the location of the subbody point, so our model of Earth rotation, and our station location, needs to be good to this level – which requires 1 s accuracy in the timing of the data.

Two types of ephemerides are available: analytic ephemerides, which provide a closed-form algebraic description of the motion of the body, and the much more precise numerical ephemerides, computed from numerical integration of the equations of motion, with parameters chosen to best fit astronomical data. While numerical ephemerides are more accurate, they are less convenient for most users, being available only as tables; analytic ephemerides are or can be made available as computer code.

The first tidal-computation program based directly on an ephemeris was that of Longman (1959), still in use for making rough tidal corrections for gravity surveys. Longman's program, like some others, computed accelerations directly, making it less obvious that an ephemeris-based approach could be applied to all tidal computations. Munk and Cartwright (1966) applied the ephemeris method to find the tidal potential. Subsequent programs (Broucke et al., 1972; Harrison, 1971; Merriam, 1992; Tamura, 1982) have used ever more precise ephemerides, with the lunar part derived from subsets of Brown's lunar theory.

Numerical ephemerides have been used primarily to produce reference time series, rather than for general-purpose programs, although the current IERS standards use such a method for computing tidal potentials and displacements (with corrections described in **Section 3.06.3.2.2**). Most precise calculations (e.g., Hartmann and Wenzel, 1995) have relied on the numerical ephemerides produced by JPL (Standish et al., 1992). The resulting tidal series forms the basis for a harmonic expansion of the tidal potential, a standard method to which we now turn.

3.06.2.3 Computing the Tides (II): Harmonic Decompositions

Since the work of Thomson and Darwin in the 1870s and 1880s, the commonest method of analyzing and predicting the tides, and of expressing tidal behavior, has been through a *harmonic expansion* of the tidal potential. In this, we express the T_{nm} as a sum of sinusoids, whose frequencies are related to combinations of astronomical frequencies and whose amplitudes are determined from the expressions in the ephemerides for R, θ', and ϕ'. In such an expansion, we write the complex T_{nm}'s as

$$T_{nm}(t) = \sum_{k=1}^{K_{nm}} A_{knm} e^{i(2\pi f_{knm} t + \varphi_{knm})} \qquad [11]$$

where, for each degree and order, we sum K_{nm} sinusoids with specified real amplitudes, frequencies, and phases A, f, and φ. The individual sinusoids are called *tidal harmonics* (not the same as the spherical harmonics of 3.1.)

This method has the conceptual advantage of decoupling the tidal potential from the details of astronomy and the practical advantage that a table of harmonic amplitudes and frequencies, once produced, is valid over a long time. Such an expansion also implicitly puts the description into the frequency domain, something as useful here as it is in other parts of geophysics. We can use the same frequencies for any tidal phenomenon, provided that it comes from a linear response to the driving potential – which is essentially true for the Earth tides. So, while this expansion was first used for ocean tides (for which it remains the standard), it works just as well for any type of Earth tides.

To get the flavor of this approach, and also introduce some terminology, we consider tides for a very simple situation: a body moving at constant speed β in a circular orbit, the orbital plane being inclined at an angle I to the Earth's equator. The angular distance from the ascending node (where the orbit plane and the equatorial plane intersect) is βt. The rotation of the Earth, at rate Ω, causes the terrestrial longitude of the ascending node to be Ωt; since the ascending node is fixed in space, Ω corresponds to one revolution per sidereal day. We further assume that at $t=0$ the body is at the ascending node and longitude 0° is under it. Finally, we take just the real part of eqn [6] and do not worry about signs.

With these simplifications we consider first the diurnal degree-2 tides ($n=2,\ m=1$). After some tedious spherical trigonometry and algebra, we find that

$$V/g = K_2\left(\frac{6\pi}{5}\right)\left[\sin I\cos I\sin\Omega t + \frac{1}{2}\sin I(1+\cos I)\sin(\Omega-2\beta)t + \frac{1}{2}\sin I(1-\cos I)\sin(\Omega+2\beta)t\right]$$

This shows that the harmonic decomposition includes three harmonics, with arguments (of time) Ω, $\Omega-2\beta$, and $\Omega+2\beta$; their amplitudes depend on I, the inclination of the orbital plane. If I were zero, there would be no diurnal tides at all. For our simple model, a reasonable value of I is 23.44°, the inclination of the Sun's orbital plane and the mean inclination of the Moon's. These numbers produce the harmonics given in **Table 2**, in which the frequencies are given in cycles per solar day (cpd). Both the Moon and Sun produce a harmonic at 1 cycle per sidereal day. For the Moon, β corresponds to a period of 27.32 days (the tropical month) and for the Sun 365.242 days (1 year), so the other harmonics are at ±2 cycles per month, or ±2 cycles per year, from this. Note that there is not a harmonic at 1 cycle per lunar (or solar) day – this is not unexpected given the degree-2 nature of the tidal potential.

Table 2 Tidal constituents (simple model)

Argument	Moon			Sun		
	Freq. (cpd)	*Period*	*Amp. (m)*	*Freq. (cpd)*	*Period*	*Amp. (m)*
Long-period tides						
	0.000000	∞	0.217	0.000000	∞	0.100
2β	0.073202	$13^d15^h51^m36^s$	0.066	0.005476	$182^d14^h45^m40^s$	0.030
Diurnal tides						
Ω	1.002738	$23^h56^m04.1^s$	0.254	1.002738	$23^h56^m04.1^s$	0.117
$\Omega-2\beta$	0.929536	$25^h49^m09.6^s$	0.265	0.997262	$24^h03^m57.2^s$	0.122
$\Omega+2\beta$	1.075940	$22^h18^m21.9^s$	0.011	1.008214	$23^h48^m16.1^s$	0.005
Semidiurnal tides						
2Ω	2.005476	$11^h58^m02.0^s$	0.055	2.005476	$11^h58^m02.0^s$	0.025
$2\Omega-2\beta$	1.932274	$12^h25^m14.2^s$	0.640	2.000000	$12^h00^m00.0^s$	0.294

It is convenient to have a shorthand way of referring to these harmonics; unfortunately, the standard naming system, now totally entrenched, was begun by Thomson for a few tides and then extended by Darwin in a somewhat *ad hoc* manner. The result is a series of conventional names that simply have to be learned as is (though only the ones for the largest tides are really important). For the Moon, the three harmonics have the *Darwin symbols* K_1, O_1, and OO_1; for the Sun, they are K_1 (again, since this has the same frequency for any body), P_1, and ϕ_1.For the semidiurnal ($m=2$ case), the result is

$$V/g = K_2\left(\frac{24\pi}{5}\right)\left[(1-\cos^2 I)\cos 2\Omega t + \frac{1}{2}(1+\cos I)^2\cos(2\Omega-2\beta)t + \frac{1}{2}(1-\cos I)^2\cos(2\Omega+2\beta)t\right]$$

again giving have three harmonics, though for I equal to 23.44°, the third one is very small. Ignoring the last term, we have two harmonics, also listed in Table 2. The Darwin symbol for the first argument is K_2; again, this frequency is the same for the Sun and the Moon, so these combine to make a lunisolar tide. The second argument gives the largest tides: for the Moon, M_2 (for the Moon) or S_2 (for the Sun), at precisely 2 cycles per lunar (or solar) day, respectively.

Finally, the $m=0$, or long-period, case has

$$V/g = K_2\left(\frac{\pi}{5}\right)\left[(1.5\sin^2 I-1)-1.5\sin^2 I\cos 2\beta t\right]$$

which gives one harmonic at zero frequency (the so-called permanent tide) and another with an argument of 2β, making tides with frequencies of two cycles per month (Mf, the fortnightly tide, from the Moon) and two cycles per year (Ssa, the semiannual tide, from the Sun).

This simple model demonstrates one other important attribute of the tides, namely, that they depend on the orbital inclination I. For the Sun, this inclination (called the obliquity of the ecliptic) changes gradually by only $1.3\times10^{-4\circ}$/year, but for the Moon, it varies by $\pm5.13^\circ$ from the mean, with a period of 18.61 years. This produces a variation in amplitude in all the lunar tides, which is called the *nodal modulation*. The simple expressions show that the resulting variation is $\pm18\%$ for O_1 and $\pm3\%$ for M_2, though the M_2 change can be larger in amplitude. Such a modulated sinusoid can be written as $\cos\omega_0 t(1+A\cos\omega_m t)$, with $\omega_0 \gg \omega_m$; this is equal to

$$\cos\omega_0 t + \frac{1}{2}A\cos[(\omega_0+\omega_m)t] + \frac{1}{2}A\cos[(\omega_0-\omega_m)t]$$

so we can retain a development purely in terms of sinusoids, but with three harmonics, one at the central frequency and two smaller ones (called *satellite harmonics*) separated from it by one cycle in 18.61 years.

An accurate ephemeris also includes the ellipticity of the orbits and all the periodic variations in I and in other orbital parameters, which lead to many more harmonics; for a detailed description, see Bartels (1957/1985). The first full harmonic expansion, including satellite harmonics, was by Doodson (1921), done algebraically from an analytic ephemeris; the result had 378 harmonics. Doodson needed a nomenclature for these tides and introduced one that relies on the fact that, as our simple ephemeris suggests, the frequency of any harmonic is the sum of multiples of a few basic frequencies. For any (n, m), we can write the argument of the exponent in eqn [11] as

$$2\pi f_k t + \phi_k = \left(\sum_{l=1}^{6} D_{lk}\, 2\pi f_l\right)t + \sum_{l=1}^{6} D_{lk}\varphi_l$$

where the f_l's are the frequencies corresponding to various astronomical periods and the φ_l's are the phases of these at some suitable epoch. Table 3 gives a list. (Recent tabulations extend this notation with up to five more arguments to describe the motions of the planets. As the tides from these are small, we ignore them here.) The $l=1$ frequency is chosen to be one cycle per lunar day exactly, so for the tide, the D_l's are 2,0,0,0,0,0. This makes the solar tide, S_2, have the D_l's 2,2,2,0,0,0. In practice, all but the smallest tides have D_{lk} ranging from -5 to 5 for $l>1$. Doodson therefore added 5 to these numbers to make a compact code, so that M_2 becomes 255·555 and $S_2$273·555. This is called the *Doodson number*; the numbers without 5 added are sometimes called Cartwright–Tayler codes (Table 4).

Figure 3 shows the full spectrum of amplitude coefficients, from the expansion of Hartmann and Wenzel (1995). The top panel shows all harmonics on a linear scale, making it clear that only a few are large, and the separation into different species around 0, 1, and 2 cycles per day: These are referred to as the *long-period*, *diurnal*, and *semidiurnal* tidal bands. The two lower panels show an expanded view of the diurnal and semidiurnal bands, using a log scale of amplitude to include the smaller harmonics. What is apparent from these is that

Table 3 Fundamental tidal frequencies

l	*Symbol*	*Frequency (cycles per day)*	*Period*	*What*
1	τ	0.9661368	$24^h50^m28.3^s$	Lunar day
2	s	0.0366011	27.3216^d	Moon's longitude: tropical month
3	h	0.0027379	365.2422^d	Sun's longitude: solar year
4	p	0.0003095	8.847^y	Lunar perigee
5	N'	0.0001471	18.613^y	Lunar node
6	p_s	0.0000001	$20\,941^y$	Solar perigee

'Longitude' refers to celestial longitude, measured along the ecliptic.

Table 4 Largest tidal harmonics, for $n=2$, sorted by size for each species

Amplitude (m)	*Doodson number*	*Frequency (cpd)*	*Darwin symbol*
Long-period tides			
−0.31459	055·555	0.0000000	M_0, S_0
−0.06661	075·555	0.0732022	Mf
−0.03518	065·455	0.0362916	Mm
−0.03099	057·555	0.0054758	Ssa
0.02793	055·565	0.0001471	N
−0.02762	075·565	0.0733493	
−0.01275	085·455	0.1094938	Mtm
−0.00673	063·655	0.0314347	MSm
−0.00584	073·555	0.0677264	MSf
−0.00529	085·465	0.1096409	
Diurnal tides			
0.36864	165·555	1.0027379	K_1
−0.26223	145·555	0.9295357	O_1
−0.12199	163·555	0.9972621	P_1
−0.05021	135·655	0.8932441	Q_1
0.05003	165·565	1.0028850	
−0.04947	145·545	0.9293886	
0.02062	175·455	1.0390296	J_1
0.02061	155·655	0.9664463	M_1
0.01128	185·555	1.0759401	OO_1
−0.00953	137·455	0.8981010	ρ_1
−0.00947	135·645	0.8930970	
−0.00801	127·555	0.8618093	σ_1
0.00741	155·455	0.9658274	
−0.00730	165·545	1.0025908	
0.00723	185·565	1.0760872	
−0.00713	162·556	0.9945243	π_1

each tidal species is split into a set of bands, separated by one cycle per month; these are referred to as *groups*: in each group, the first two digits of the Doodson number are the same. All harmonics with the same first three digits of the Doodson number are in clusters separated by one cycle per year; these clusters are called *constituents*, though this name is also sometimes used for the individual harmonics. As a practical matter, this is usually the finest frequency resolution attainable; on the scale of this plot, finer frequency separations, such as the nodal modulation, are visible only as a thickening of some of the lines. All this fine-scale structure poses a challenge to tidal analysis methods (**Section 3.06.5.1**).

Since Doodson provided the tidal potential to more than adequate accuracy for studying ocean tides, further developments did not take place for the next 50 years, until Cartwright and Tayler (1971) revisited the subject. Using eqn [6], they computed the potential from a more modern lunar ephemeris and then applied special Fourier methods to analyze, numerically, the resulting series and get amplitudes for the various harmonics. The result was a compendium of 505 harmonics, which (with errors corrected by Cartwright and Edden, 1973) soon became the standard under the usual name of the CTE representation. (A few small harmonics at the edges of each band, included by Doodson but omitted by Cartwright, are sometimes added to make a CTED list with 524 harmonics.)

More extensive computations of the tidal potential and its harmonic decomposition have been driven by the very high precision available from the ephemerides and the desire for more precision for analyzing some tidal data (gravity tides from superconducting gravimeters). Particular expansions are those of Bullesfeld (1985), Tamura (1987), Xi (1987), Hartmann and Wenzel (1995), and Roosbeek (1995). The latest is that of Kudryavtsev (2004), with 27 000 harmonics. **Figure 4** shows the amplitude versus number of harmonics; to get very high accuracy demands a very large number. But not many are needed for a close approximation; the CTED expansion is good to about 0.1% of the total tide.

3.06.2.4 The Pole Tide

Both in our elementary discussion and in our mathematical development of the tidal forcing, we treated the Earth's rotation only as a source of motion of the subbody point. But changes in this rotation also cause spatial variations in the gravitational potential, and since these have the same effects as the attraction of external bodies, they can also be regarded as tides. The only significant one is the *pole tide*, which is caused by changes in the direction of the Earth's spin axis relative to a reference frame fixed in the Earth. The spin produces a centrifugal force, which depends on the angular distance between the spin axis and some location: so as the spin axis moves, this distance, and the centrifugal force, changes.

For a spin vector Ω, the potential at a location r is

$$V = \frac{1}{2}\left[|\Omega|^2|r|^2 - |\Omega\cdot r|^2\right] \qquad [12]$$

We assume that the rotation vector is nearly along the 3-axis, so that we have $\Omega = \Omega(m_1\hat{x}_1 + m_2\hat{x}_2 + \hat{x}_3)$, with m_1 and m_2 both much less than one; m_1 and m_2 describe the instantaneous rotation axis. If we put expression for this axis into eqn [12], and subtract V for m_1 and m_2 both zero, we get the potential height for the pole tide:

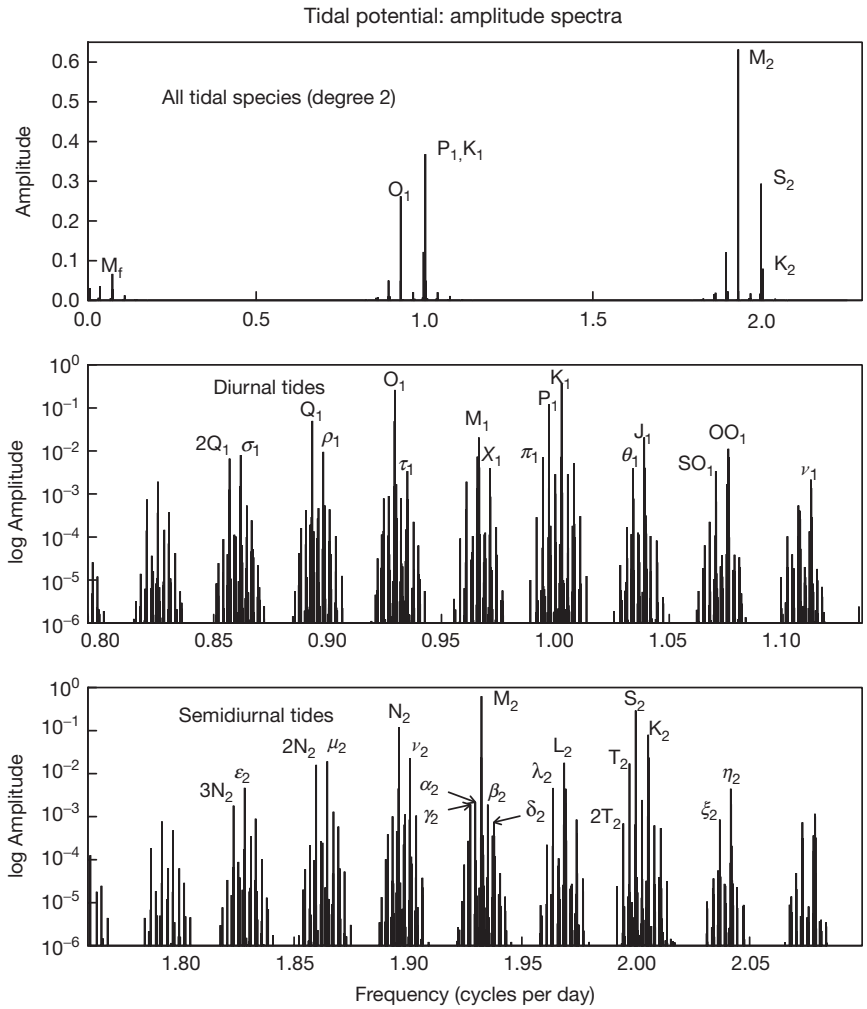


Figure 3 The spectrum of the tidal potential. Since all variations are purely sinusoidal, the spectrum is given by the amplitudes of the tidal harmonics, taken from Hartmann and Wenzel (1995), though normalized according to the convention of Cartwright and Tayler (1971). The Darwin symbols are shown for the larger harmonics (top) and all named diurnal and semidiurnal harmonics, except for a few that are shown only in **Figure 5**.

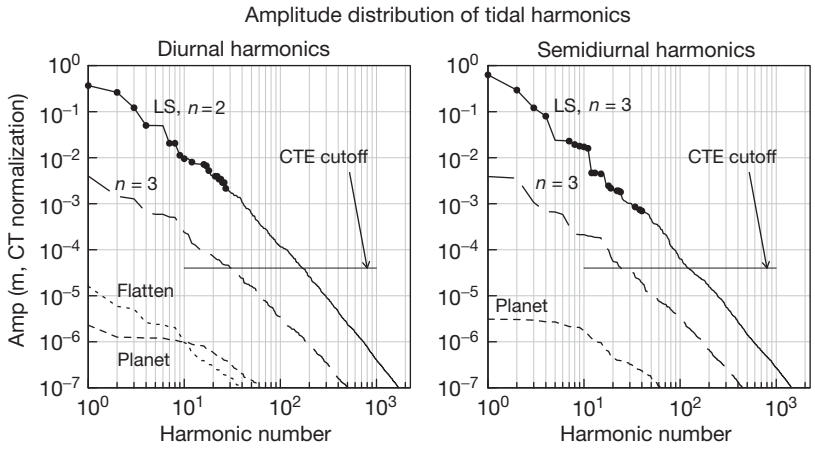


Figure 4 Distribution of harmonic amplitudes for the catalog of Hartmann and Wenzel (1995), normalized according to Cartwright and Tayler (1971). The line 'LS, $n=2$' refers to lunisolar harmonics of degree two; those with large dots have Darwin symbols associated with them. Constituents of degree 3, from other planets, and from Earth flattening are shown as separate distributions. The horizontal line shows the approximate cutoff level of the Cartwright and Tayler (1971) list.

$$\frac{V}{g} = -\frac{\Omega^2}{2g}[2(m_1 r_1 r_3 + m_2 r_2 r_3)]$$
$$= -\frac{\Omega^2 a^2}{g}\sin\theta\cos\theta(m_1\cos\phi + m_2\sin\phi)$$

This is a degree-2 change in the potential, of the same form as for the diurnal tides. However, the periods involved are very different, since the largest pole tides come from the largest polar motions, at periods of 14 months (the Chandler wobble) and 1 year. The maximum range of potential height is a few cm, small but not negligible; pole-tide signals have been observed in very precise gravity measurements and also, not strongly, in sea-level data (Desai, 2002; Haubrich and Munk, 1959; Miller and Wunsch, 1973; Trupin and Wahr, 1990). Like the polar motion, the actual pole tide can only be found from observations of the Earth's orientation. These do not actually give the position of $\boldsymbol{\Omega}$, but rather that of what is called the celestial intermediate pole (CIP), whose coordinates are denoted by x_p and y_p (Pétit and Luzum, 2010). The position of $\boldsymbol{\Omega}$ is given by

$$m_1 = x_p - \bar{x}_p - \frac{\dot{y}_p}{\Omega} \quad m_2 = -y_p + \bar{y}_p - \frac{\dot{x}_p}{\Omega} \qquad [13]$$

where $\bar{x}, \bar{y}$ are mean values of the pole position (Pétit and Luzum, 2010, Section 7.1.4) and the third term of each expression corrects for the difference between the CIP and $\boldsymbol{\Omega}$ (Brzezínski and Capitaine, 1993; Gross, 1992); at periods of a year and longer, this term is much smaller than the others.

3.06.2.5 Radiational Tides

The harmonic treatment used for the gravitational tides can also be applied to the various phenomena associated with solar heating. As a first approximation, the input radiation can be taken to be proportional to the cosine of the Sun's elevation during the day and zero at night; Munk and Cartwright (1966) called this the *radiational tide*. The day–night asymmetry produces harmonics of degrees 1 and 2; these have been tabulated by Cartwright and Tayler (1971) and are shown in **Figure 5** as crosses (for both degrees), along with the tidal potential harmonics shown as in **Figure 3**. The unit for the radiational tides is S, the solar constant: 1366 W m^{-2}.

Changes in solar irradiation drive changes in air temperature and pressure in very complicated ways but changes in ground temperature fairly directly. Ground-temperature changes in turn cause thermoelastic deformations with these periods (Berger, 1975; Harrison and Herbst, 1977; Müller, 1977). Air pressure changes driven by solar heating, usually known as *atmospheric tides* (Chapman and Lindzen, 1970), load the Earth enough to cause deformations and changes in gravity. The availability of better models of some atmospheric tides (Ray, 2001; Ray and Ponte, 2003; Ray and Poulose, 2005) and their inclusion in ocean-tide models (Ray and Egbert, 2004) has allowed their effects to be compared with gravity observations (Boy et al., 2006b).

That some of these thermal tidal harmonics coincide with harmonics in the tidal potential poses a real difficulty for precise analysis of the latter, since if all we have is the sum of two harmonics with the same frequency, it is impossible to tell how much each part contributes. The only way to estimate this is by making additional assumptions, about how the response to the potential or to temperature behaves at other frequencies. Even then, it is likely that these estimates will have large systematic errors; this is why when estimating tidal responses, the large K_1 tide is less used than the smaller O_1 tide.

3.06.3 Tidal Response of the Solid Earth

Having described the tidal forces, we next turn to the response of the solid Earth – which, as is conventional, we assume to be oceanless, putting in the effect of the ocean tides at a later step. We start with the usual approximation of a spherical Earth in order to introduce a number of concepts, many of them adequate for all but the most precise modeling of the tides. We then describe what effects a better approximation has, in enough detail to enable computation; the relevant theory is beyond the scope of this treatment, though outlined in the article by Dehant and Mathews.

3.06.3.1 Tidal Response of an SNREI Earth

To a good approximation, we can model the tidal response of an oceanless Earth by assuming an SNREI Earth model: that is, one that is *s*pherical, *n*on-*r*otating, *e*lastic, and *i*sotropic. As in normal-mode seismology (from which this acronym comes), this means that we assume the elastic properties to vary only with depth. In addition to these restrictions on the Earth

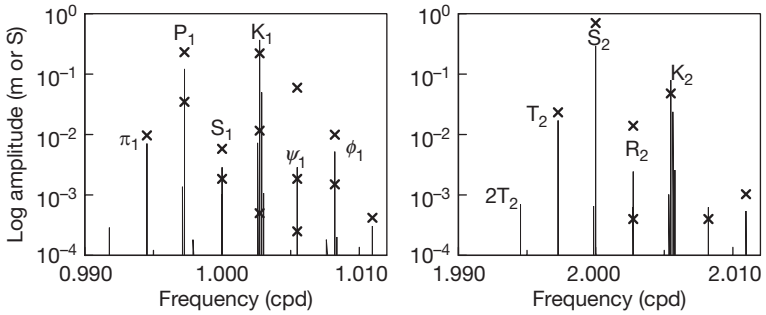


Figure 5 Radiational tides. The crosses show the amplitudes of the radiational tidal harmonics (degree 1 and 2) from Cartwright and Tayler (1971); the amplitudes are for the solar constant S being taken to be 1.0. The lines are harmonics of the gravitational tides, as in **Figure 3**.

model, we add one more about the tidal forcing: that it has a much longer period than any normal modes of oscillation of the Earth so that we can use a quasistatic theory, taking the response to be an equilibrium one. Since the longest-period normal modes for such an Earth have periods of less than an hour, this is a good approximation.

It is simple to describe the response of an SNREI Earth to the tidal potential (Jeffreys, 1976). Because of symmetry, only the degree n is relevant. If the potential height at a point on the surface is $V(\theta,\phi)/g$, the distortion of the Earth from tidal forces produces an additional gravitational potential $k_nV(\theta,\phi)$, a vertical (i.e., radial) displacement $h_nV(\theta,\phi)/g$, and a horizontal displacement $l_n(\nabla_1V(\theta,\phi)/g)$, where ∇_1 is the horizontal gradient operator on the sphere. So defined, k_n, h_n, and l_n are dimensionless; they are called *Love numbers*, after A. E. H. Love (though the parameter l_n was actually introduced by T. Shida). For a standard modern Earth model (PREM), $h_2=0.6032$, $k_2=0.2980$, and $l_2=0.0839$. For comparison, the values for the much older Gutenberg–Bullen Earth model are 0.6114, 0.3040, and 0.0832 – not very different. In this section, we adopt values for a and g that correspond to a spherical Earth: 6.3707×10^6 m and 9.821 m s$^{(-2)}$, respectively.

3.06.3.1.1 Some combinations of Love numbers (I): gravity and tilt

Until there were data from space geodesy, neither the potential nor the displacements could be measured; what could be were ocean tides, tilt, changes in gravity, and local deformation (strain), each of which possessed its own expression in terms of Love numbers – which we now derive. Since the first three of these would exist even on a rigid Earth, it is common to describe them using the ratio between what they are on an elastic Earth (or on the real Earth) and what they would be on a rigid Earth.

The simplest case is that of the effective tide-raising potential: that is, the one relevant to the ocean tide. The total tide-raising potential height is $(1+k_n)V/g$, but the solid Earth (on which a tide gauge sits) goes up by h_nV/g, so the effective tide-raising potential is $(1+k_n-h_n)V/g$, sometimes written as γ_nV/g, γ_n being called the *diminishing factor*. For the PREM model, $\gamma_2=0.6948$. Since tilt is just the change in slope of an equipotential surface, again relative to the deforming solid Earth, it scales in the same way that the potential does: the tilt on an SNREI Earth is γ_n times the tilt on a rigid Earth. The NS tilt is, using eqn [8] and expressions for the derivatives of Legendre functions,

$$\lambda_N=\frac{-\gamma_n}{ga}\frac{\partial V}{\partial\theta}=\frac{-1}{a\sin\theta}\sum_{n=2}^{n=\infty}\gamma_n\sum_{m=0}^{n}N_n^m\left[n\cos\theta P_n^m(\cos\theta)-(n+m)P_{n-1}^m(\cos\theta)\right]\left[a_n^m(t)\cos m\phi+b_n^m(t)\sin m\phi\right] \quad [14]$$

where the sign is chosen such that a positive tilt to the north would cause a plumb line to move in that direction. The east tilt is

$$\lambda_E=\frac{\gamma_n}{ga\sin\theta}\frac{\partial V}{\partial\phi}=\frac{-1}{a\sin\theta}\sum_{n=2}^{n=\infty}\gamma_n\sum_{m=0}^{n}mN_n^mP_n^m(\cos\theta)\left[b_n^m(t)\cos m\phi-a_n^m(t)\sin m\phi\right] \quad [15]$$

with the different combinations of a and b with the ϕ dependence showing that this tilt is phase-shifted relative to the potential, by 90° if we use a harmonic decomposition.

Tidal variations in gravity were for a long time the commonest type of Earth-tide data. For a spherical Earth, the tidal potential is, for degree n,

$$V_n\left(\frac{r}{a}\right)^n+k_nV_n\left(\frac{a}{r}\right)^{n+1}$$

where the first term is the potential caused by the tidal forcing (and for which we have absorbed all nonradial dependence into V_n) and the second is the additional potential induced by the Earth's deformation. The corresponding change in local gravitational acceleration is the radial derivative of the potential:

$$\frac{\partial}{\partial r}\left[V_n\left(\left(\frac{r}{a}\right)^n+k_n\left(\frac{a}{r}\right)^{n+1}\right)\right]_{r=a}=V_n\left[\frac{n}{a}-(n+1)\frac{k_n}{a}\right] \quad [16]$$

In addition to this change in gravity from the change in the potential, there is a change from the gravimeter being moved up by an amount h_nV_n/g. The change in gravity is this displacement times the gradient of g, $2g/a$, plus the displacement times $-\omega^2$, where ω is the radian frequency of the tidal motion – that is, the inertial acceleration. (We adopt the Earth-tide convention that a decrease in g is positive.) If we ignore this last part (which is at most 1.5% of the gravity-gradient part), we get a total change of

$$V_n\left[\frac{n}{a}-\left(\frac{n+1}{a}\right)k_n+\frac{2h_n}{a}\right]=\frac{nV_n}{a}\left[1-\left(\frac{n+1}{n}\right)k_n+\frac{2}{n}h_n\right] \quad [17]$$

The nV_n/a term is the change in g that would be observed on a rigid Earth (with h and k zero); the term which this is multiplied by, namely,

$$\delta_n=1+\frac{2h_n}{n}-\left(\frac{n+1}{n}\right)k_n$$

is called the *gravimetric factor*. For the PREM model, $\delta_2=1.1562$: the gravity tides are only about 16% larger than they would be on a completely rigid Earth, so that most of the tidal gravity signal shows only that the Moon and Sun exist, but does not provide any information about the Earth. The expression for the gravity tide is of course very similar to eqn [8]:

$$\delta g=\frac{g}{a}\sum_{n=2}^{n=\infty}\delta_n\sum_{m=0}^{n}N_n^mP_n^m(\cos\theta)\left[a_n^m(t)\cos m\phi+b_n^m(t)\sin m\phi\right] \quad [18]$$

3.06.3.1.2 Combinations of Love numbers (II): Displacement and strain tides

For a tidal potential of degree n, the displacements at the surface of the Earth ($r=a$) will be, by the definitions of the Love numbers l_n and h_n,

$$u_r=\frac{h_nV}{g}\quad u_\theta=\frac{l_n}{g}\frac{\partial V}{\partial\theta}\quad u_\phi=\frac{l_n}{g\sin\theta}\frac{\partial V}{\partial\phi} \quad [19]$$

in spherical coordinates. Comparing these with [18], [14], and [15], we see that the vertical displacement is exactly proportional to changes in gravity, with the scaling constant being $h_na/2\delta_ng=1.692\times10^5$ s^2, and that the horizontal displacements are exactly proportional to tilts, with the scaling

constant being $l_n a/\gamma_n = 7.692 \times 10^5$ m; we can thus use eqns [18], [14], and [15], suitably scaled, to find tidal displacements.

Taking the derivatives of [19], we find the tensor components of the surface strain are

$$e_{\theta\theta} = \frac{1}{ga}\left(h_n V + l_n \frac{\partial^2 V}{\partial^2 \theta}\right)$$

$$e_{\phi\phi} = \frac{1}{ga}\left(h_n V + l_n \cot\theta \frac{\partial V}{\partial \theta} + \frac{l_n}{\sin\theta}\frac{\partial^2 V}{\partial^2 \phi}\right)$$

$$e_{\theta\phi} = \frac{l_n}{ga\sin\theta}\left(\frac{\partial^2 V}{\partial\theta\partial\phi} - \cot\theta\frac{\partial V}{\partial\phi}\right)$$

We again use [8] for the tidal potential and get the following expressions that give the formulas for the three components of surface strain for a particular n and m; to compute the total strain, these should be summed over all $n \geq 2$ and all m from 0 to n (though in practice, the strain tides with $n > 3$ or $m = 0$ are unobservable):

$$e_{\theta\theta} = \frac{N_n^m}{a\sin^2\theta}\left[\left(h_n\sin^2\theta + l_n\left(n^2\cos^2\theta - n\right)\right)P_n^m(\cos\theta) - 2l_n(n-1)\right.$$
$$\left.(n+m)\cos\theta P_{n-1}^m(\cos\theta) + l_n(n+m)(n+m-1)P_{n-2}^m(\cos\theta)\right]\cdot$$
$$\left[a_n^m(t)\cos m\phi + b_n^m(t)\sin m\phi\right]$$

$$e_{\phi\phi} = \frac{N_n^m}{a\sin^2\theta}\left[\left(h_n\sin^2\theta + l_n\left(n\cos^2\theta - m^2\right)\right)P_n^m(\cos\theta)\right.$$
$$\left.-l_n(n+m)\cos\theta P_{n-1}^m(\cos\theta)\right]\cdot\left[a_n^m(t)\cos m\phi + b_n^m(t)\sin m\phi\right]$$

$$e_{\theta\phi} = \frac{mN_n^m l_n}{a\sin^2\theta}\left[(n-1)\cos\theta P_n^m(\cos\theta) - (n+m)P_{n-1}^m(\cos\theta)\right]\cdot$$
$$\left[b_n^m(t)\cos m\phi - a_n^m(t)\sin m\phi\right]$$

Note that the combination of the longitude factors with $a_n^m(t)$ and $b_n^m(t)$ means that $e_{\theta\theta}$ and $e_{\phi\phi}$ are in phase with the potential, while $e_{\theta\phi}$ is not.

One consequence of these expressions is that, for n equal to two, and m equal to either one or two, the areal strain, $(1/2)(e_{\theta\theta} + e_{\phi\phi})$, is equal to $V(h_2 - 3l_2)/ga$: the areal strain, vertical displacement, potential, and gravity are all scaled versions of each other. Close to the surface, the free-surface condition means that deformation is nearly that of plane stress, so vertical and volume strains are also proportional to areal strain and are also just a scaled version of the potential.

If we combine these expressions for spatial variation with the known amplitudes of the tidal forces, we can see how the rms amplitude of the body tides varies with latitude and direction (**Figure 6**). There are some complications in the latitude dependence; for example, the EW semidiurnal strain tides go to zero at 52.4° latitude. Note that while the tilt tides are larger than strain tides, most of this signal is from the direct attractions of the sun and moon; the purely deformational part of the tilt is about the same size as the strain.

3.06.3.2 Response of a Rotating Earth

We now turn to models for tides on an oceanless and isotropic Earth, still with properties that depend on depth only but add rotation and slightly inelastic behavior. Such models have three consequences for the tidal response:

1. The ellipticity of the core–mantle boundary and the rotation of the Earth combine to produce a free oscillation in which the fluid core (restrained by pressure forces) and solid mantle precess around each other. This is known as the nearly diurnal free wobble (NDFW) or free core nutation. Its frequency falls within the band of the diurnal tides, which causes a resonant response in the Love numbers near one cycle per day. The diurnal tides also cause changes in the direction of the Earth's spin axis (the astronomical precessions and nutations), and the NDFW affects these as well.
2. Ellipticity and rotation couple the response to forcing of degree n to spherical harmonics of other degrees and spheroidal to toroidal modes of deformation. As a result, the Love numbers become slightly latitude-dependent, and additional terms appear for horizontal displacement.
3. The imperfect elasticity of the mantle (finite Q) modifies the Love numbers in two ways: they become complex, with small imaginary parts; and they become weakly frequency-dependent because of anelastic dispersion.

The full theory for these effects, especially the first two, is quite complicated. Love (1911) provided some theory for the effects of ellipticity and rotation, and Jeffreys and Vincente (1957) and Molodenskii (1961) for the NDFW, but the modern approach for these theories was first described by Wahr (1981a,b); a simplified version is given by Neuberg et al. (1987) and Zürn (1997b). For more recent developments, see the article by Dehant and Mathews in this volume and Mathews et al. (1995a,b, 1997), Wang (1997), Dehant et al. (1999), Mathews et al. (2002), and Mathews and Guo (2005).

To obtain the full accuracy of these theories, particularly for the NDFW correction, requires the use of tabulated values of the Love numbers for specific harmonics, but analytic approximations are also available. The next three sections outline these, using values from the IERS standards (Pétit and Luzum, 2010) for the Love numbers and from Dehant et al. (1999) for the gravimetric factor.

3.06.3.2.1 NDFW resonance

The most important result to come out of the combination of improved theoretical development and observations has been that the frequency of the NDFW, f_{NFDW}, is significantly different from that originally predicted. This frequency is usually expressed by the period of the NDFW in a space-fixed frame, $T_{\mathrm{NDFW}} = (f_{\mathrm{NFDW}} - 1.0027379)^{-1}$, where f_{NFDW} is the resonance frequency for an Earth-fixed frame, in cycles per solar day. T_{NDFW} is in large part determined by the ellipticity of the core–mantle boundary (CMB), for a hydrostatic Earth theory gives $T_{\mathrm{NDFW}} = 460$ d. Both tidal and nutation data (Amoruso et al., 2012; Herring et al., 2002; Rosat and Lambert, 2009; Zürn, 1997a) give values of about 430 d. This difference implies that the ellipticity of the CMB departs from a hydrostatic value by about 5%, the equivalent of a 500 m variation difference in radius. This departure is generally thought to reflect distortion of the core–mantle boundary by mantle convection (e.g., Steinberger and Holme, 2008).

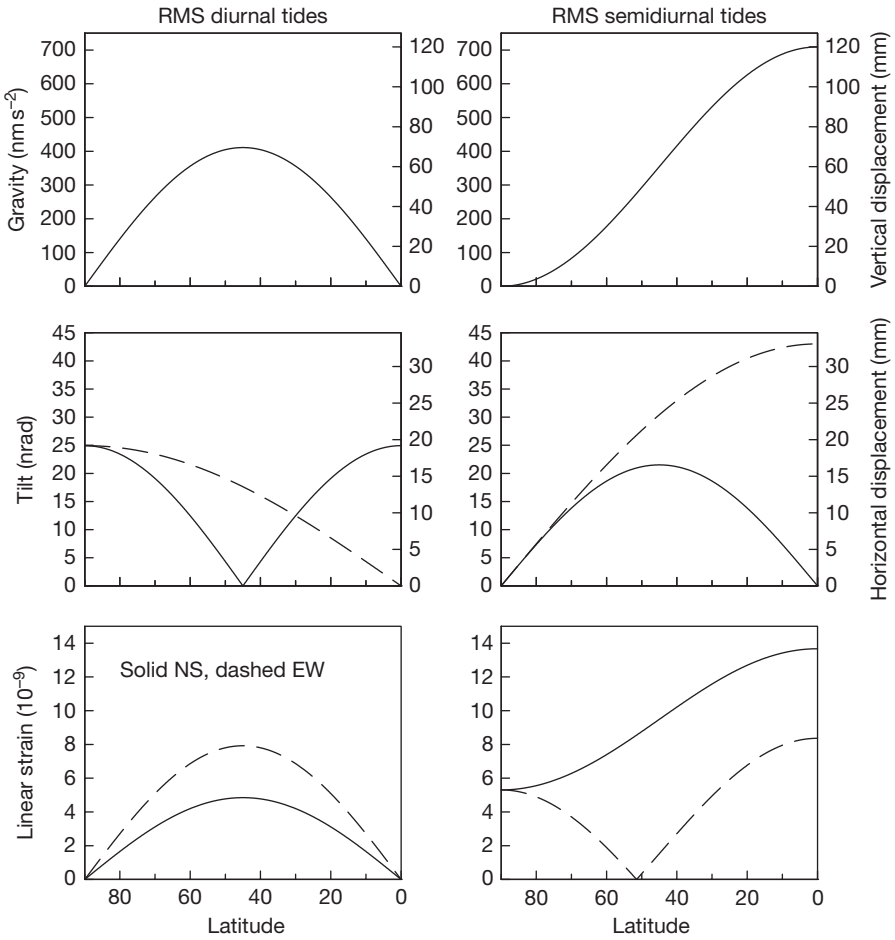


Figure 6 RMS tides. The left plots show the rms tides in the diurnal band, and the right plots the rms in the semidiurnal band. The uppermost frame shows gravity and vertical displacement (with scales for each), which have the same latitude dependence; the next horizontal displacement and tilt; and the bottom linear strain. In all plots but the top, dashed is for EW measurements, solid for NS.

Table 5 Coefficients (real and imaginary parts) used in eqn [20] to find the frequency dependence of the Love numbers (including corrections for ellipticity) in the diurnal tidal band

	$\Re(S_z)$	$\Im(S_z)$	$\Re(S_1)$	$\Im(S_1)$	$\Re(S_2)$	$\Im(S_2)$
δ_0	1.15802	0.0	-2.871×10^{-3}	0.0	4.732×10^{-5}	0.0
$k^{(0)}$	0.29954	-1.412×10^{-3}	-7.811×10^{-4}	-3.721×10^{-5}	9.121×10^{-5}	-2.971×10^{-6}
$h^{(0)}$	0.60671	-2.420×10^{-3}	-1.582×10^{-3}	-7.651×10^{-5}	1.810×10^{-4}	-6.309×10^{-6}
$l^{(0)}$	0.08496	-7.395×10^{-4}	-2.217×10^{-4}	-9.672×10^{-6}	-5.486×10^{-6}	-2.998×10^{-7}
δ_+	1.270×10^{-4}	0.0	-2.364×10^{-5}	0.0	1.564×10^{-6}	0.0
k^+	-8.040×10^{-4}	2.370×10^{-6}	2.090×10^{-6}	1.030×10^{-7}	-1.820×10^{-7}	6.500×10^{-9}
$h^{(2)}$	-6.150×10^{-4}	-1.220×10^{-5}	1.604×10^{-6}	1.163×10^{-7}	2.016×10^{-7}	2.798×10^{-9}
$l^{(2)}$	1.933×10^{-4}	-3.819×10^{-6}	-5.047×10^{-7}	-1.643×10^{-8}	-6.664×10^{-9}	5.090×10^{-10}
$l^{(1)}$	1.210×10^{-3}	1.360×10^{-7}	-3.169×10^{-6}	-1.665×10^{-7}	2.727×10^{-7}	-8.603×10^{-9}
l^P	-2.210×10^{-4}	-4.740×10^{-8}	5.776×10^{-7}	3.038×10^{-8}	1.284×10^{-7}	-3.790×10^{-9}

The resonant behavior of the Love numbers from the NDFW is confined to the diurnal band; within that band, it can be approximated by an expansion in terms of the resonant frequencies:

$$L(f) = S_z + \sum_{k=1}^{2} \frac{S_k}{f - f_k} \qquad [20]$$

where $L(f)$ is the frequency-dependent Love number (of whatever type) for frequency f in cycles per solar day. The expansion in use for the IERS standards includes three resonances: the Chandler wobble, the NDFW, and the free inner core nutation (FICN); to a good approximation (better than 1%), the last can be ignored. Table 5 gives the values of the S's according to the IERS standards (and to Dehant et al., 1999 for the gravimetric

factors), scaled for f in cycles per solar day; in these units, the resonance frequencies are

$$f_1 = -2.60812 \times 10^{-3} - 1.365 \times 10^{-4} i$$
$$f_2 = 1.0050624 + 2.5 \times 10^{-5} i$$

and the FICN frequency (not used in eqn [20]) is $1.00176124 + 7.82 \times 10^{-4} i$. Dehant et al. (1999) used purely real-valued frequencies, with $f_1 = -2.492 \times 10^{-3}$ and $f_2 = 1.0050623$, as well as purely real values of the S's.

Figure 7 shows the NDFW resonance, for a signal (areal strain) relatively sensitive to it. Unfortunately, the tidal harmonics do not sample the resonance very well; and because the largest effect is for the small ψ_1 harmonic, which is also affected by radiational tides, the best estimates of f_{NDFW} have come from nutation rather than from tidal data.

One consequence of the NDFW resonance is that we cannot use equations of the form [18] to compute the theoretical diurnal tides, since the factor for them varies with frequency. If we construct the $a_n^m(t)$ and $b_n^m(t)$ using [11], it is easy to adjust the harmonic amplitudes and phases appropriately. Alternatively, if we find $a_n^m(t)$ and $b_n^m(t)$ using an ephemeris, we can compute the diurnal tides assuming a frequency-independent factor and then apply corrections for the few harmonics that are both large and affected by the resonance; Mathews et al. (1997) and Pétit and Luzum (2010) described such a procedure for displacements and for the induced potential.

3.06.3.2.2 Coupling to other modes

The other effect of rotation and ellipticity is to couple the spheroidal deformation of degree n, driven by the potential, to spheroidal deformations of degree $n \pm 2$ and toroidal deformations of degree $n \pm 1$. Thus, the response to the degree-2 part of the tidal potential contains a small component of degrees 0 and 4. If we generalize the Love numbers, describing the response as a ratio between the response and the potential, the result will be a ratio that depends on latitude: we say that the Love number has become latitude-dependent.

Such a generalization raises issues of normalization; unlike the spherical case, the potential and the response may be evaluated on different surfaces. This has been a source of some confusion. The normalization of Mathews et al. (1995b) is the one generally used for displacements: it uses the response in displacement on the surface of the ellipsoid but takes these to be relative to the potential evaluated on a sphere with the Earth's equatorial radius. Because of the inclusion of such effects as the inertial and Coriolis forces, the gravimetric factor is no longer the combination of potential and displacement Love numbers, but an independent ratio, defined as the ratio of changes in gravity on the ellipsoid to the direct attraction at the same point. Both quantities are evaluated along the normal to the ellipsoid, as a good approximation to the local vertical. Wahr (1981a) used the radius vector instead, producing a much larger apparent latitude effect.

The standard expression for the gravimetric factor is given by Dehant et al. (1999):

$$\delta(\theta) = \delta_0 + \delta_+ \frac{Y_{n+2}^m}{Y_n^m} + \delta_- \frac{Y_{n-2}^m}{Y_n^m} \qquad [21]$$

By the definition of the Y_n^m's, $\delta_- = 0$ except for $m \le n-2$, so for the $n = 2$ tides, we have for the diurnal tides

$$\delta(\theta) = \delta_0 + \delta_+ \frac{\sqrt{3}}{2\sqrt{2}} \left(7\cos^2\theta - 3\right)$$

and for the semidiurnal tides,

$$\delta(\theta) = \delta_0 + \delta_+ \frac{\sqrt{3}}{2} \left(7\cos^2\theta - 1\right)$$

The expression for the induced potential (Wahr, 1981a) is similar, namely, that the potential is obtained by replacing the term $Y_{nm}(\theta, \phi)$ in equation [7] with

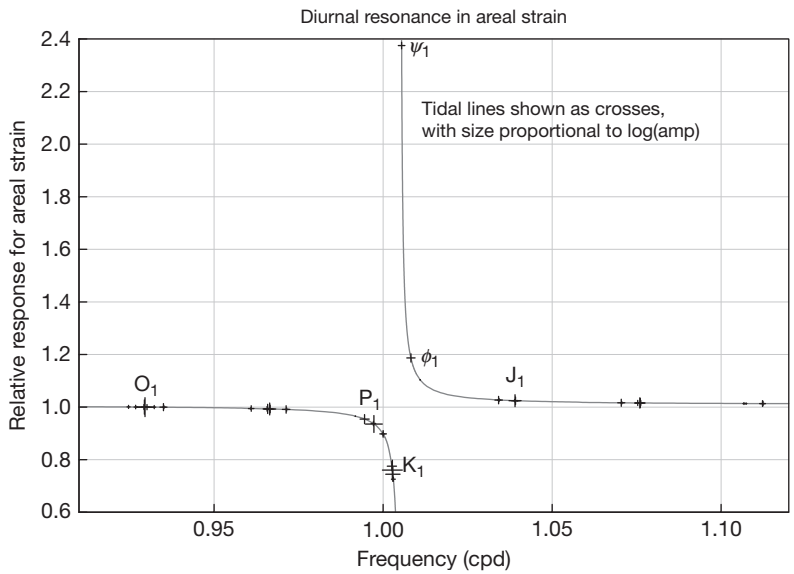


Figure 7 NDFW response for the combination of Love numbers that gives the areal strain, normalized to 1 for the O1 tide. The crosses show the locations of tidal harmonics, with size of symbol proportional to the logarithm of the amplitude of the harmonic.

$$k_0\left(\frac{a_e}{r}\right)^{n+1}Y_{nm}(\theta,\phi)+k_+\left(\frac{a_e}{r}\right)^{n+3}Y^m_{n+2}(\theta,\phi) \qquad [22]$$

which recovers the conventional Love number for $k_+=0$.

The expressions for displacements are more complicated, not only because this is a vector quantity but also because the horizontal displacements include spheroidal–toroidal coupling, which affects neither the vertical, the potential, nor the gravity. For the degree-2 tides, the effect of coupling to the degree-4 deformation is allowed for by defining

$$h(\theta)=h^{(0)}+\frac{1}{2}h^{(2)}\left(3\cos^2-1\right)$$
$$l(\theta)=l^{(0)}+\frac{1}{2}l^{(2)}\left(3\cos^2\theta-1\right) \qquad [23]$$

Then, to get the vertical displacement, replace the term $Y_{nm}(\theta,\phi)$ in equation [7] with

$$h(\theta)Y^m_2(\theta,\phi)+\frac{\delta_{m0}h^P}{N^0_2} \qquad [24]$$

where the δ_{m0} is the Kronecker delta, since h^P (usually called h' in the literature) only applies for $m=0$; the N^0_2 factor arises from the way in which h^P was defined by Mathews et al. (1995b).

The displacement in the $\hat{\theta}$ (north) direction is obtained by replacing the term $Y_{nm}(\theta,\phi)$ in eqn [7] with

$$l(\theta)\frac{\partial Y^m_2(\theta,\phi)}{\partial\theta}-\frac{ml^1\cos\theta}{\sin\theta}Y^m_2(\theta,\phi)+\frac{\delta_{m1}l^P}{N^1_2}e^{i\phi} \qquad [25]$$

where again the l^P (usually called l') applies only for the particular value of $m=1$, for which it applies a correction such that there is no net rotation of the Earth. And finally, to get the displacement in the $\hat{\phi}$ (east) direction, we replace the $Y_{nm}(\theta,\phi)$ term in eqn [7] with

$$i\left[\frac{ml(\theta)}{\sin\theta}Y^m_2(\theta,\phi)+l^1\cos\theta\frac{\partial Y^m_2(\theta,\phi)}{\partial\theta}+\frac{\delta_{m0}\sin\theta l^P}{N^0_2}\right] \qquad [26]$$

where again the l^P term applies, for $m=0$, a no-net-rotation correction. The multiplication by i means that when this is applied to [7] and the real part taken, the time dependence will be $b^m_n(t)\cos m\phi-a^m_n(t)\sin m\phi$ instead of $a^m_n(t)\cos m\phi+b^m_n(t)\sin m\phi$.

Table 6 gives the generalized Love numbers for selected tides, including ellipticity, rotation, the NDFW, and anelasticity (which we discuss later). The values for the diurnal tides are from exact computations rather than the resonance approximations given earlier. It is evident that the latitude dependence ranges from small to extremely small, the latter applying to the gravimetric factor, which varies by only 4×10^{-4} from the equator to 60° N. For the displacements, Mathews et al. (1997) showed that the various coupling effects are at most 1 mm; the latitude dependence of $h(\theta)$ changes the predicted displacements by 0.4 mm out of 300.

3.06.3.2.3 Anelastic effects

All modifications to the Love numbers discussed so far apply to an Earth model that is perfectly elastic. However, the materials of the real Earth are slightly dissipative (anelastic), with a finite Q. Measurements of the Q of Earth tides were long of interest because of their possible relevance to the problem of tidal evolution of the Earth–Moon system (Cartwright, 1999); though it is now clear that almost all of the dissipation of tidal energy occurs in the oceans (Ray et al., 2001), anelastic effects on tides remain of interest because tidal data (along with the Chandler wobble) provide the only information on Q at frequencies below about 10^{-3} Hz.

Over the seismic band (approximately 10^{-3} to 1 Hz), Q appears to be approximately independent of frequency. A general model for frequency dependence is

$$Q=Q_0\left(\frac{f}{f_0}\right)^{\alpha} \qquad [27]$$

where f_0 and Q_0 are reference values. In general, in a dissipative material, the elastic modulus μ will in general also be a function of frequency, with $\mu(f)$ and $Q(f)$ connected by the Kramers–Kronig relation (Dahlen and Tromp, 1998, **Chapter 1.07**). (We use μ because this usually denotes the shear modulus; in pure compression, Q is very high and the Earth can be treated as elastic.) This frequency dependence is usually termed *anelastic dispersion*. For the frequency dependence of Q given by eqn [27], and α small, the modulus varies as

Table 6 Love numbers for an Earth that includes ellipticity, rotation, anelasticity, and the NDFW resonance

	Ssa	*Mf*	O_1	P_1	K_1	ψ_1	M_2	M_3
δ_0	1.15884	1.15767	1.15424	1.14915	1.13489	1.26977	1.16172	1.07338
δ_+	0.00013	0.00013	0.00008	−0.00010	−0.00057	0.00388	0.00010	0.00006
δ_-	−0.00119	−0.00118						
$\Re[k^{(0)}]$	0.3059	0.3017	0.2975	0.2869	0.2575	0.5262	0.3010	0.093
$\Im[k^{(0)}]$	−0.0032	−0.0021	−0.0014	−0.0007	0.0012	0.0021	−0.0013	
k^+	−0.0009	−0.0009	−0.0008	−0.0008	−0.0007	−0.0011	−0.0006	
$\Re[h^{(0)}]$	0.6182	0.6109	0.6028	0.5817	0.5236	1.0569	0.6078	0.2920
$\Im[h^{(0)}]$	−0.0054	−0.0037	−0.0006	−0.0006	−0.0006	−0.0020	−0.0022	
$\Re[l^{(0)}]$	0.0886	0.0864	0.0846	0.0853	0.0870	0.0710	0.0847	0.0015
$\Im[l^{(0)}]$	−0.0016	−0.0011	−0.0006	−0.0006	−0.0007	−0.0001	−0.0006	
$h^{(2)}$	−0.0006	−0.0006	−0.0006	−0.0006	−0.0007	−0.0001	−0.0006	
$l^{(2)}$	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	
$l^{(1)}$			0.0012	0.0012	0.0011	0.0019	0.0024	
l^P			−0.0002	−0.0002	−0.0003	−0.0001		

Values for the gravimetric factors are from Dehant et al. (1999) and for the other Love numbers from the IERS standards (Pétit and Luzum, 2010).

$$\mu(f) = \mu_0 \left[1 + \frac{1}{Q_0} \left\{ \frac{2}{\alpha\pi} \left[1 - \left(\frac{f_0}{f} \right)^{\alpha} \right] + i \left(\frac{f_0}{f} \right)^{\alpha} \right\} \right] \quad [28]$$

so there is a slight variation in the modulus with frequency, and the modulus becomes complex, introducing a phase lag into its response to sinusoidal forcing. In the limit as α approaches zero (constant Q), the real part has a logarithmic frequency dependence. Including a power-law variation [27] at frequencies below a constant-Q seismic band, the frequency dependence becomes

$$\mu(f) = \mu_0 \left[1 + \frac{1}{Q_0} \left\{ \frac{2}{\pi} \ln \left(\frac{f_0}{f} \right)^{\alpha} + i \right\} \right] \quad f > f_m \quad [29]$$

$$\mu(f) = \mu_0 \left[1 + \frac{1}{Q_0} \left\{ \frac{2}{\alpha\pi} \left[\alpha \ln \left(\frac{f_m}{f_0} \right) + 1 - \left(\frac{f_m}{f} \right)^{\alpha} \right] + i \left(\frac{f_m}{f} \right)^{\alpha} \right\} \right] \quad f < f_m$$

where f_m is the frequency of transition between the two Q models.

Adding anelasticity to an Earth model has three effects on the computed Love numbers:

1. Anelastic dispersion means that the elastic constants of an Earth model found from seismology must be adjusted slightly to be appropriate for tidal frequencies. As an example, Dehant et al. (1999) found that for an elastic Earth model, the gravimetric factor δ_0 is 1.16030 for the M_2 tide; an anelastic model gives 1.16172; for $h^{(0)}$, the corresponding values are 0.60175 and 0.61042.
2. Dispersion also means that the Love numbers vary within the tidal bands. For the semidiurnal and diurnal tides, the effect is small, especially compared to the NDFW resonance; in the long-period bands, it is significant as f approaches zero. The usual formulation for this (Pétit and Luzum, 2010) is based on a slightly different form of [29], from Smith and Dahlen (1981) and Wahr and Bergen (1986); the Love numbers vary in the long-period band as

$$L(f) = A - B \left\{ \cot \frac{\alpha\pi}{2} \left[1 - \left(\frac{f_m}{f} \right)^{\alpha} \right] + i \left(\frac{f_m}{f} \right)^{\alpha} \right\} \quad [30]$$

   where A and B are constants for each Love number. For the IERS standards, $\alpha = 0.15$ and $f_m = 432$ cpd (a period of 200 s). A and B are 0.29525 and -5.796×10^{-4} for k^0, 0.5998 and -9.96×10^{-4} for $h^{(0)}$, and 0.0831 and -3.01×10^{-4} for $l^{(0)}$.
3. As eqn [30] shows, the Love numbers also become complex-valued, introducing small phase lags into the tides. There are additional causes for this; in particular, the NDFW frequency can have a small imaginary part because of dissipative core–mantle coupling, and this will produce complex-valued Love numbers even in an elastic Earth. Complex-valued Love numbers can be used in extensions of eqns [7] and [8]; for example, if the elastic Love number combination introduces no phase shift (as for gravity) in phase, the real part is multiplied by $[a_n^m(t) \cos m\phi + b_n^m(t) \sin m\phi]$ and the imaginary part by $[b_n^m(t) \cos m\phi - a_n^m(t) \sin m\phi]$.

The most recent examination of tidal data for anelastic effects (Benjamin et al., 2006) combined data from diurnal tides (in the potential, as measured by satellites), the Chandler wobble, and the 19-year nodal tide. They find a good fit for α between 0.2 and 0.3 with $f_m = 26.7$ cpd (see also Ray and Egbert, 2012); using the IERS value of f_m gave a better fit for α between 0.15 and 0.25.

3.06.3.3 Lateral Heterogeneities

Relative to an SNREI Earth, the Earth's ellipticity can be viewed as a lateral heterogeneity with very long wavelength. Since the Earth is heterogeneous on scales shorter than this, a natural question is what effect these lateral heterogeneities will have on the Earth tides. Because of the range of spatial scales for lateral heterogeneities, this is a very broad question, ranging from differences over thousands of kilometers to variations with scales of less than a meter close to the measurements. The last are usually referred to as site effects and are discussed separately in **Section 3.06.6.1**.

One approach (Fu and Sun, 2007; Molodensky, 1977) is to express the lateral heterogeneities as spherical harmonics and use perturbation theory to compute their effects. Other methods include numerical procedures such as spectral elements (Metivier et al., 2005, 2006) or finite-volume methods (Latychev et al., 2009). All of these have been applied to find the effect of lateral heterogeneities on gravity and displacement. Given realistic lateral heterogeneities in density, elastic moduli, and prestress, the perturbations to the semidiurnal tide are less than 1 mm for the radial displacements for radial motion and 2 nm s^{-2} for gravity. Comparing with **Figure 6** shows that this is no more than 5×10^{-3} in displacement and 10^{-3} in gravity. While gravity measurements have a very high precision, in comparing them with theory, the limiting factor is their accuracy: as described in **Section 3.06.6**, gravimeter calibrations are at best good to only 10^{-3}. It may be easier to observe perturbations in displacements – and widespread GPS measurements provide many more estimates of these than are available from tidal gravimeters.

3.06.4 Tidal Loading

A major barrier to using Earth tides to find out about the solid Earth is the observations also include the effects of the ocean tides – which, depending on what is being studied, may be signal or noise. Even on a rigid Earth, the redistribution of mass in the ocean tides would create signals from the attraction of the water; on the real Earth, this redistribution also causes the Earth to deform, which creates additional changes. These induced signals are called the *load tides*, which combine with the body tide to make up the total tide (**Figure 1**).

3.06.4.1 Computing Loads I: Spherical Harmonic Sums

To compute the load, we start with a description of the ocean tides, almost always as a complex-valued function $H(\theta', \phi')$, giving the amplitude and phase of a particular constituent over the ocean; we discuss such ocean-tide models in more detail in **Section 3.06.4.4**. The loads can then be computed in two ways: using a sum of spherical harmonics or as a convolution of the tide height with a Green function.

In the first approach, we expand the tidal elevation in spherical harmonics:

$$H(\theta', \phi') = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} H_{nm} Y_{nm}(\theta', \phi') \quad [31]$$

where Y_{nm} are as in the section on tidal forcing and H_{nm} would be found from

$$H_{nm} = \int_0^{\pi} \sin\theta' \, \mathrm{d}\theta' \int_0^{2\pi} \mathrm{d}\phi H(\theta', \phi') Y_{nm}^* \equiv \int_{\Omega} H(\theta', \phi') Y_{nm}^* \, \mathrm{d}\Omega \quad [32]$$

where we use Ω for the surface of the sphere. Note that there will be significant high-order spherical harmonic terms in H_{nm}, because the tidal height goes to zero over land: any function with a step behavior will decay only gradually with increasing degree.

The mass distribution H causes a gravitational potential on the surface of the Earth, which we call V^L. This potential is given by the integral over the surface of the potential function times H; the potential function is proportional to r^{-1}, where r is the linear distance from the location (θ, ϕ) to the mass at (θ', ϕ'), making the integral

$$V^L(\theta, \phi) = G\rho_w a^2 \int_{\Omega} \frac{H(\theta', \phi')}{r} \mathrm{d}\Omega \quad [33]$$

where ρ_w is the density of seawater and G and a are as in Section 3.06.2.1. We can write the r^{-1} in terms of angular distance Δ:

$$\begin{aligned} \frac{1}{r} &= \frac{1}{2a\sin(\Delta/2)} = \frac{1}{a} \sum_{n=0}^{\infty} P_n(\cos\Delta) \\ &= \frac{1}{a} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{4\pi}{2n+1} Y_{nm}(\theta', \phi') Y_{nm}^*(\theta, \phi) \end{aligned} \quad [34]$$

where we have again used the addition theorem [5]. Combining the last expression in eqn [34] with the spherical harmonic expansion [31] and the expression for the potential [33] gives the potential in terms of spherical harmonics:

$$V^L = G\rho_w a \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{4\pi}{2n+1} H_{nm} Y_{nm}(\theta, \phi) \quad [35]$$

We have found the potential produced by the load because this potential is used, like the tidal potential, in the specification of the Earth's response to the load. Specifically, we define the *load Love numbers* k_n', h_n', and l_n' such that, for a potential V^L of degree n, we have

$$u_n^z = h_n' \frac{V_n^L}{g} \quad u_n^h = l_n' \frac{\nabla_1 V_n^L}{g} \quad V_n = k_n' V_n^L \quad [36]$$

where u_n^z is the vertical displacement (also of degree n), u_n^h is the horizontal displacement, and V_n is the additional potential produced by the deformation of the Earth. These load Love numbers, like the Love numbers for the tidal potential, are found by integrating the differential equations for the deformation of the Earth, but with a different boundary condition at the surface: a normal stress from the load, rather than zero stress. For a spherical Earth, these load numbers depend only on the degree n of the spherical harmonic.

To compute the loads, we combine the definition of the load Love numbers [36] with the expression [35], using whichever combination is appropriate for some observable. For example, for vertical displacement u^z, this procedure gives

$$\begin{aligned} u^z(\theta, \phi) &= \frac{G\rho_w a}{g} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{4\pi}{2n+1} h_n' H_{nm} Y_{nm}(\theta, \phi) \\ &= \frac{\rho_w}{\rho_E} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{3h_n'}{2n+1} H_{nm} Y_{nm}(\theta, \phi) \end{aligned} \quad [37]$$

where ρ_E is the mean density of the Earth. A similar expression applies for the induced potential, with k_n' replacing h_n'; for the effective tide-raising potential, sometimes called the 'self-attraction loading' or SAL (Ray, 1998), we would use $1 + k_n' - h_n'$.

Many terms are needed for a sum in eqn [37] to converge, but such a sum provides the response over the whole Earth. Ray and Sanchez (1989) used this method to compute radial displacement over the whole Earth, with $n = 256$, and special methods to speed the computation of the H_{nm} coefficients in eqn [32]. In any method that sums harmonics, there is always room for concern about the effects of Gibbs' phenomenon (Hewitt and Hewitt, 1979) near discontinuities, but no such effect was observed in the displacements computed near coastlines. Mitrovica et al. (1994) independently developed the same method, extended it to the more complicated case of horizontal displacements, and were able to make calculations with $n = 2048$.

Given a global ocean-tide model and a need to find loads over the entire surface, this summation technique requires much less computation than the convolution methods to be discussed in the next section. For gravity, the contributions from the Earth's response are, from eqn [17], $-(n+1)k_n' V_n^L/a$ from the induced potential and $2h_n' V_n^L/a$ from the displacement, making the sum

$$\delta g(\theta, \phi) = \frac{3g\rho_w}{a\rho_E} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{2h_n' - (n+1)k_n'}{2n+1} H_{nm} Y_{nm}(\theta, \phi) \quad [38]$$

While this sum might appear to converge more slowly than eqn [37] because of the $n+1$ multiplying k_n', the convergence is similar because for large n, nk_n' approaches a constant value, which we term k_∞'. All three load Love numbers have such asymptotic limits for large n:

$$\text{As} \quad n \Rightarrow \infty \quad h_n' \Rightarrow h_\infty' \quad nk_n' \Rightarrow k_\infty' \quad nl_n' \Rightarrow l_\infty' \quad [39]$$

so that the sum [38] converges reasonably well. A similar sum can be used to get the gravity from the direct attraction of the water:

$$\frac{3g\rho_w}{a\rho_E} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{4n+2} H_{nm} Y_{nm}(\theta, \phi)$$

(Agnew, 1983; Merriam, 1980).

However, summation over harmonics is not well suited to quantities that involve spatial derivatives, such as tilt or strain. To find the load tides for these, we need instead to employ convolution methods, which we now turn to.

3.06.4.2 Computing Loads II: Integration Using Green Functions

If we only want the loads at a few places, the most efficient approach is to multiply the tide model by a *Green function*,

which gives the response to a point load, and integrate over the area that is loaded by the tides. That is, we work in the spatial domain rather than, as in Section 3.06.4.1, in the wave number domain; there is a strict analogy with Fourier theory, in which convolution of two functions is the same as multiplying their Fourier transforms. Convolution methods have other advantages, such as the ability to combine different ocean-tide models easily, include more detail close to the point of observation, and handle any of the Earth-tide observables. The standard reference on the procedure remains the classic paper of Farrell (1972); Jentzsch (1997) is a more recent summary.

More formally, we find the integral over the sphere (in practice over the oceans)

$$\int_0^{\pi} r\sin\theta'\,\mathrm{d}\theta' \int_0^{2\pi} r\,\mathrm{d}\phi' G_{\mathrm{L}}(\theta,\phi,\theta',\phi')\rho_{\mathrm{w}}H(\theta',\phi') \qquad [40]$$

where G_{L} is the *Green function* for an effect (of whatever type) at (θ,ϕ) from a point mass (δ-function) at (θ',ϕ'); $\rho_{\mathrm{w}}gHr^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$ is the applied force.

The Green functions are found, not directly, but by forming sums of combinations of the load Love numbers. The first step is to find the potential from a point mass. Take $H=\rho_{\mathrm{w}}a^2\delta(\theta',\phi')$ (where δ is the Dirac delta function). Substitute this into eqn [33], using the sum in $P_n(\cos\Delta)$ in eqn [34]. This gives the potential as

$$V^L(\theta,\phi)=G\rho_{\mathrm{w}}a\int_{\Omega}H(\theta',\phi')\sum_{n=0}^{\infty}P_n(\cos\Delta)\mathrm{d}\Omega=\frac{ga}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}P_n(\cos\Delta) \qquad [41]$$

which shows that the degree-n part of the potential is $V_n^L=ga/M_{\mathrm{E}}$, independent of n. So, to compute vertical displacement, we would apply this potential to the load Love number h_n', getting the displacement

$$u^z=\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}h_n'\frac{V_n^L}{g}=\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}h_n'P_n(\cos\Delta)=G_z(\Delta) \qquad [42]$$

which is thus the loading Green function for vertical displacement. Some insight into the behavior of this function can be obtained by using the asymptotic value of h_n' to write

$$\begin{aligned}G_z(\Delta)&=\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}h_\infty'P_n(\cos\Delta)+\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}(h_n'-h_\infty')P_n(\cos\Delta)\\&=\frac{ah_\infty'}{2M_{\mathrm{E}}\sin\Delta/2}+\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}(h_n'-h_\infty')P_n(\cos\Delta)\end{aligned} \qquad [43]$$

where we have made use of [34]. The new sum will converge much more rapidly; in practice, it needs to include only enough terms for h_n' to have approached h_∞' to adequate precision. For Δ small, the sum approaches zero, so the analytic part shows that, for loads nearby, G_z varies as Δ^{-1}. This is the vertical displacement seen for a point load on an elastic half-space, in what is called the Boussinesq solution; in the limit of short distance, the loading problem reduces to this, which provides a useful check on numerical computations.

For the horizontal displacement, the Green function is

$$u^h=\frac{ga}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}\frac{l_n'}{g}\frac{\partial V_n^L}{\partial\Delta}=\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}l_n'\frac{\partial P_n(\cos\Delta)}{\partial\Delta}=G_{\mathrm{h}}(\Delta) \qquad [44]$$

which may again have the asymptotic part nl_n' removed and replaced by an analytic expression

$$\begin{aligned}G_{\mathrm{h}}(\Delta)&=\frac{al_\infty'}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}\frac{1}{n}\frac{\partial P_n(\cos\Delta)}{\partial\Delta}+\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}(l_n'-l_\infty')\frac{\partial P_n(\cos\Delta)}{\partial\Delta}\\&=-\frac{al_\infty'}{M_{\mathrm{E}}}\frac{\cos(\Delta/2)[1+2\sin\Delta/2]}{2\sin(\Delta/2)[1+\sin\Delta/2]}\\&\quad+\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}(nl_n'-l_\infty')\frac{1}{n}\frac{\partial P_n(\cos\Delta)}{\partial\Delta}\end{aligned} \qquad [45]$$

which shows the same dependence on Δ for small distances.

For gravity, there are two parts to the loading: the direct attraction of the water mass (often called the Newtonian part) and the change caused by elastic deformation of the Earth. The first part can be found analytically by using the inverse square law and computing the vertical part of the attraction. If the elevation of our point of observation is εa, with ε small, this Green function is

$$G_{\mathrm{gn}}(\Delta)=-\frac{g}{M_{\mathrm{E}}}\left[\frac{\varepsilon+2\sin^2\Delta/2}{(4(1+\varepsilon)\sin^2\Delta/2+\varepsilon^2)^3/2}\right] \qquad [46]$$

The elastic part of the Green function follows from the harmonic expression [38]:

$$\begin{aligned}G_{\mathrm{ge}}=\frac{g}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}(2h_n'-(n+1)k_n')P_n(\cos\Delta)=\frac{g}{M_{\mathrm{E}}}\Big[&2(h_n'-h_\infty')\\&+\frac{h_\infty'}{\sin\Delta/2}-k_\infty'\left(\frac{1}{2\sin\Delta/2}-1-\log\left(\sin\Delta/2+\sin^2\Delta/2\right)\right)\\&+\sum_{n=1}^{\infty}\left(2(h_n'-h_\infty')-(n+1)\left(k_n'-\frac{k_\infty'}{n}\right)\right)P_n(\cos\Delta)\Big]\end{aligned} \qquad [47]$$

which shows, again, a Δ^{-1} singularity for Δ small.Likewise, the Green function for the tide-raising potential makes use of the combination $1-k_n'+h_n'$:

$$\begin{aligned}G_{\mathrm{pot}}&=\frac{a}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}(1+k_n'-h_n')P_n(\cos\Delta)\\&=\frac{a(1-h_\infty')}{2M_{\mathrm{E}}\sin\Delta/2}+\sum_{n=0}^{\infty}(k_n'-(h_n'-h_\infty'))P_n(\cos\Delta)\end{aligned} \qquad [48]$$

and the Green function for tilt uses the same combination of Love numbers, but with the derivative of the P_n's:

$$\begin{aligned}G_{\mathrm{t}}&=\frac{-1}{M_{\mathrm{E}}}\sum_{n=0}^{\infty}(1+k_n'-h_n')\frac{\partial P_n(\cos\Delta)}{\partial\Delta}\\&=\frac{(1-h_\infty')\cos(\Delta/2)}{4M_{\mathrm{E}}\sin^2\Delta/2}-\sum_{n=0}^{\infty}(k_n'-(h_n'-h_\infty'))\frac{\partial P_n(\cos\Delta)}{\partial\Delta}\end{aligned} \qquad [49]$$

which has a Δ^{-2} singularity for Δ small. The tilt is thus much more sensitive to local loads than the other observables we have so far discussed.

The remaining Green function is that for strain, specifically for the strain in the direction of the load

$$e_{\Delta\Delta}=\frac{1}{a}\frac{\partial u^h}{\partial\Delta}+\frac{u^z}{a}$$

from which the Green function is, from [42] and [44],

$$G_{\Delta\Delta} = \frac{1}{M_E}\sum_{n=0}^{\infty} h'_n P_n(\cos\Delta) + \frac{1}{M_E}\sum_{n=0}^{\infty} l'_n \frac{\partial^2 P_n(\cos\Delta)}{\partial\Delta^2}$$
$$= \frac{h'_\infty}{2M_E \sin\Delta/2} + \frac{1}{M_E}\sum_{n=0}^{\infty}(h'_n - h'_\infty)P_n(\cos\Delta)$$
$$+ \frac{l'_\infty}{M_E}\sum_{n=0}^{\infty}\frac{1}{n}\frac{\partial^2 P_n(\cos\Delta)}{\partial\Delta^2} + \frac{1}{M_E}\sum_{n=0}^{\infty}(l'_n - l'_\infty)\frac{\partial^2 P_n(\cos\Delta)}{\partial\Delta^2}$$
$$= \frac{h'_\infty}{2M_E \sin\Delta/2} + \frac{1}{M_E}\sum_{n=0}^{\infty}(h'_n - h'_\infty)P_n(\cos\Delta)$$
$$+ \frac{l'_\infty}{M_E}\frac{1+\sin\Delta/2+\sin^2\Delta/2}{4\sin^2\Delta/2[1+\sin\Delta/2]} + \frac{1}{M_E}\sum_{n=0}^{\infty}(l'_n - l'_\infty)\frac{\partial^2 P_n(\cos\Delta)}{\partial\Delta^2} \quad [50]$$

which again shows a near-field singularity of Δ^{-2}. This behavior also holds for the strain perpendicular to the direction to the load; since this is given by

$$\frac{u^z}{a} + \frac{\cos\Delta}{\sin\Delta}\frac{u^h}{a} \quad [51]$$

there is no need to compute a separate Green function for it. The Green function for linear strain that would be used in [40] is

$$G_L(\Delta,\zeta) = G_{\Delta\Delta}(\Delta)\cos^2\zeta + \left[\frac{G_z(\Delta)}{a} + \cot\Delta\frac{G_h(\Delta)}{a}\right]\sin^2\zeta \quad [52]$$

where ζ is the azimuth of the load relative to the direction of extension. Areal strain has a complicated dependence on distance, because it is zero for a point load on a half-space, except right at the load.

All of the Green functions are computed by finding the load Love numbers for a range of n and forming the various sums. Farrell (1972) formed the sums up to $n = 10\,000$; the Love numbers can be computed at values of n spaced logarithmically and interpolated to intermediate values. Several numerical methods to accelerate the convergence of the sums are described by Farrell (1972), Francis and Dehant (1987), and Guo et al. (2004). The Green functions tabulated by Farrell (1972) (with the addition of the potential function by Farrell (1973)) are still widely used; Jentzsch (1997) and Guo et al. (2004) tabulated the functions for the PREM model. Kamigaichi (1998) had discussed the variations in the strain and tilt Green functions at shallow depths, forming sums up to $n = 4 \times 10^6$; his results show a smooth transition between the surface functions for an SNREI Earth model and the Boussinesq solution for a half-space, but at a nonzero depth. The response at depth lacks the singularity present in the strain and tilt Green functions at the surface. Examinations of the extent to which local structure, particularly lateral variations, affects computed load tides have not been plentiful – perhaps mostly because the data most sensitive to such effects, strain, and tilt, are affected by other local distortions (**Section 3.06.6.1**). As with the Love numbers for the body tides, the load Love numbers will be affected by rotation, ellipticity, anisotropy, and anelasticity. The first two produce, again, a resonant response from the NDFW, though only for loads of degree two and order one (Wahr and Sasao, 1981); Pagiatakis (1990) has examined effects from the others.

It is also possible to define load Love numbers, and Green functions derived for them, for transverse rather than normal stress, to describe the deformation of the Earth by wind stress or ocean currents, including tidal currents; see Merriam (1985, 1986) and Wilhelm (1986).

3.06.4.3 Degree-One Effects in Ocean Loading

Satellite-based measurements introduce a complication into ocean-loading computations, one most easily explained by supposing the Earth to be rigid. All the load Love numbers would then be zero and so would the predicted displacements from tidal loading. But observed displacements would be nonzero because the origin of a satellite reference frame is the center of mass of everything below the satellite orbit: solid Earth, ocean, and atmosphere. Ocean tides on a rigid Earth would shift this combined center of mass (CM) relative to the center of mass of the solid Earth (CE), to which the observation points would be tied. This varying displacement between CM and CE is usually termed motion of the geocenter.

In the spherical harmonic expansion of tidal height (eqn [31]), the only terms that correspond to a change in the center of mass of the ocean (and hence the displacement between CE and CM) are those with n equal to one. Because only the degree-one load Love numbers (eqn [36]), h'_1, k'_1, and l'_1, combine with the $n = 1$ terms in the load potential (eqn [35]), only these numbers change when computing loading relative to CM instead of to CE. Before space geodesy, all Earth-tide observations were made entirely on the solid Earth, and degree-one load Love numbers were computed assuming CE to be fixed.

Farrell (1972) discussed the difference between CE and CM loading briefly; most subsequent treatments assumed that this could be completely represented by a rigid-body displacement of all points. However, because the Earth deforms, a motion of CM relative to CE can involve more general motions. Blewitt (2003) showed that adding a constant to all three degree-one load Love numbers produces a motion that leaves the Earth spherical, but with patterns of vertical and horizontal displacement that depend on the constant added. The particular constant to be added to convert from CE to CM is given by considering the induced potential. If CE is fixed, the additional potential caused by the Earth's deformation is zero, which we write as $k'_{1:\mathrm{CM}} = 0$.

If we take CM as fixed, a degree-one load does not change the external potential at all, so we must have $k'_{1:\mathrm{CM}} = -1$ to cancel the direct effect of the mass causing the load. To compute loads in the CM frame, we thus need to subtract one from the degree-one load Love numbers before using them, whether in harmonic sums such as eqn [37] or in Green functions such as eqn [42]. It remains true that it is incorrect to compute CM loads for phenomena referenced to the geocenter, such as tidal gravity measured on the Earth's surface: though actually eqns [47]–[49] are all unaffected by constant shifts to the degree-one load Love numbers.

In fact, as Blewitt (2003) pointed out, the CE point cannot be determined from surface observations; but it can be approximated by choosing the load Love numbers to produce displacements whose average is minimized over the Earth's surface, a choice Blewitt terms the CF frame. For the Love numbers given by Farrell (1972), the degree-one load Love numbers for that frame differ from those for the CE frame by

amounts that cause only a 2% difference in displacements, so we can take the CF frame as equivalent to the CE frame.

Fu et al. (2012) analyzed a large suite of GPS data to compare the effects of using the CM or CE loads in processing GPS data with predetermined orbits and clocks. If the same load computation (CE or CM) was used both in finding the orbits and in reprocessing the data, the residual tides were not large; but if different load computations were used, these tides became significant. Current orbit products are all set in the CM frame, so loads computed using these Love numbers should be used for all modeling of GPS station displacements.

3.06.4.4 Ocean Tide Models

To compute loads, we also need a description of the ocean tides. Finding the global pattern of the ocean tides is a difficult task that has been pursued for some time (Cartwright, 1977, 1999). The intractability of the relevant equations, the complexity of the geometry, and the sensitivity of the results to details in the models long precluded numerical solutions, one difficulty being that the oceans have barotropic modes of oscillation with periods close to the diurnal and semidiurnal tidal bands. At the same time, it was very difficult to measure tides in deep water. All this meant that until relatively recently, there were no good tidal models for computing loads.

From the Earth-tide standpoint, what is important is that increasing computational power has rendered numerical solutions possible for realistic geometries and that satellite altimetry has provided global coverage for tidal data. Starting in the late 1990s, these developments made it possible to construct global tidal models, whose data are mostly (often entirely) derived from altimetry and computed on a relatively coarse mesh (0.125°–0.5°). Comparisons between different models (Andersen et al., 1995; Desai et al., 1997; Le Provost et al., 1995; Shum et al., 1997) showed variations in the open ocean of less than 1 cm. With more altimetry data and more computing power, additional models (Egbert and Erofeeva, 2002; Lyard et al., 2006; Matsumoto et al., 2000) have become available, and loads computed from them usually differ by only small amounts.

But there is still room for improvement. The polar regions are poorly covered by altimetry data, and much effort has gone into improving tidal models there (King and Padman, 2005; Padman and Erofeeva, 2004; Padman et al., 2002). More generally, global models often lack the resolution (in grid size and representation of bathymetry) to accurately model tides in marginal seas, many of which have large tides because of local resonances. There is thus still a need for local tidal models, to be used when Earth-tide data are collected nearby. Such models use a finer mesh and rely more on local tide-gauge data, than global models do. Ray et al. (2011) reviewed the state of local tidal modeling, which in many places remains imperfect.

Most tidal models (global or local) are given for particular tidal constituents, usually at least one diurnal and one semidiurnal. Unless a local resonance is present, the loads for other harmonics can be found by scaling using the ratios of the amplitudes in the equilibrium tide (Le Provost et al., 1991).

3.06.4.5 Computational Methods

Essentially, all load programs perform the convolution [40] directly, either over the grid of ocean cells (perhaps more finely divided near the load) or over a radial grid. Two that are generally available are GOTIC (Matsumoto et al., 2001) and SPOTL (Agnew, 1996, 2012). Bos and Baker (2005) and Penna et al. (2008) had compared the results from different programs for gravity and displacement, finding variations of a few percent because of different computational assumptions and different versions of the coastline. Most global ocean-tide models do not represent coastlines very accurately, so some local refinement is often needed. Fortunately, this can be provided using the global coastal representations made available by Wessel and Smith (1996) – except in the Antarctic, where this coastline is (in places) the ice shelves rather than the grounding line (Bindschadler et al., 2011; Rignot et al., 2011).

Figure 8 shows the computed loads for a region (Northwest Europe) with large and complex local tides. The vertical displacement and gravity loads have roughly similar forms, but the tilt and linear strain have a very different pattern, being much more concentrated near the coast, as might be expected from the different near-field behavior of their Green functions.

3.06.5 Analyzing and Predicting Earth Tides

As noted in 'Abstract,' the tidal forces can be described to extremely high precision and accuracy, and the body tides and tidal loading can often be modeled to an accuracy that exceeds that of tidal measurements. Such measurements do however provide a check on these models and in some cases allow them to be improved, so we briefly describe how tidal parameters are extracted from the data and how the data are obtained. This is important whether we aim to measure the tides, use modeled tides as a calibration signal, or predict the tides to high accuracy to check the quality of ongoing measurements.

3.06.5.1 Tidal Analysis and Prediction

As noted in 'Abstract,' the analysis of time series for tidal response is just a special case of finding the transfer function, or admittance, of a linear system, a concept first introduced into tidal analysis by Munk and Cartwright (1966). Because the tides are very bandlimited, we can find the *tidal admittance*, $W_T(f)$, only for frequencies at which $X_T(f)$ contains significant energy. For ocean tides, it is most meaningful to take $x_T(t)$ to be the local value of the tide-raising potential or for some analyses the tide computed for a nearby site (Cartwright et al., 1969). In Earth tide studies, it may be more convenient to take as reference the tides expected for an oceanless, but otherwise realistic, Earth model, so that any departure of $W(f)$ from unity will then reflect the effect of ocean loads or the inadequacy of the model.

The theory described in **Section 3.06.3.2** shows that, except for the NDFW resonance, $W(f)$ for an oceanless Earth varies only very slightly with frequency. The ocean tides show more variability, but only in limited areas do they have resonances within the tidal bands, so that in general, the ocean load also

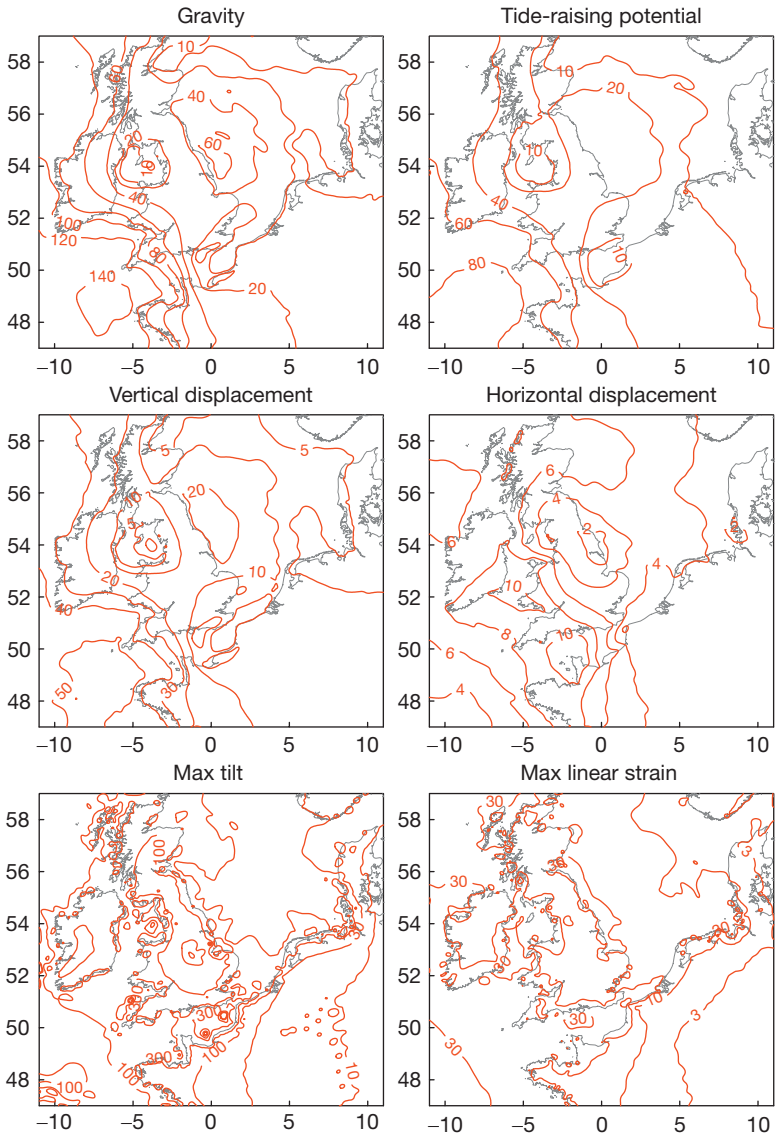


Figure 8 Loads for the M_2 tide, computed by the Green function method for a 2004 empirical global model (GOT4.7, developed using the methods of Ray et al., 1999), combined with a detailed model (done by Egbert and Erofeeva (2002) in 2010) of the tides on the European Shelf. Only the amplitude of the complex-valued quantities is shown; for tilt, displacement, and strain, the value is taken along the azimuth that maximizes the amplitude.

varies smoothly with frequency (Garrett and Munk, 1971). Even the local resonances in certain bays and gulfs have a low enough Q that the response is smoothly varying over (say) the entire semidiurnal band (Ku et al., 1985). So, the more closely spaced two frequencies are, the closer the corresponding values of $W(f)$ will be, an assumption Munk and Cartwright (1966) dubbed the 'credo of smoothness.'

A naive way to find the tidal response is to take the Fourier transform of the data (using a fast Fourier transform) and use the amplitudes and phases of the result. This is a poor choice for two reasons. One problem is that the frequencies computed by the usual definition of the discrete Fourier transform usually do not coincide with the frequencies of the tidal harmonics – especially if the length of the transform, N, is chosen to work well with a fast Fourier transform algorithm. In addition, any noise will bias the amplitudes of the coefficients to be larger than the true values.

If spectral analysis is to be used, a much better method is the cross spectral technique described by Munk and Cartwright (1966). This method has poorer frequency resolution than others to be discussed but makes the fewest assumptions about the form of $W(f)$ and also provides estimates of the noise as a function of frequency. This is useful because many methods assume the noise to be the same at all frequencies, and it may not be; in particular, the noise is sometimes observed to rise sharply in the tidal bands, a phenomenon called 'tidal cusping' (Colosi and Munk, 2006; Munk et al., 1965; Ponchaut et al., 2001). The cross spectral method does however require large amounts of data to perform reliably. The procedure is described in full by Munk and Cartwright (1966);

it depends on finding the cross spectrum between a noise-free reference series and the data, using a slow Fourier transform to make the Fourier frequencies match the tidal frequencies relatively well, windowing to reduce bias from spectral leakage, and averaging to get a statistically consistent estimate.

By far the commonest approach to tidal analysis is least-squares fitting of a set of sinusoids with known frequencies, namely, the frequencies of the largest tidal constituents. That is, we aim to minimize the sum of squares of residuals:

$$\sum_{n=0}^{N}\left[y_n-\sum_{l=1}^{L}\left(A_l\cos\left(2\pi f_l t_n\right)+B_l\sin\left(2\pi f_l t_n\right)\right)\right]^2 \qquad [53]$$

which expresses the fitting of L sine–cosine pairs with frequencies f_l to the N data y_n, the f's being fixed to the tidal harmonic frequencies and the A's and B's being solved for.

The usual assumption behind a least-squares analysis is that the residual after fitting the sinusoids will be statistically independent random variables; but this is valid only if the noise spectrum is white, which is usually not so. One departure from whiteness is the presence of increased long-period noise outside the tidal bands; this can be removed by filtering the data before analyzing it. A more difficult problem is the tidal cusping just referred to. If the noise spectrum rises to a higher level inside the tidal bands, perhaps very much higher around the frequencies of the radiational tides, this needs to be allowed for in fitting the tides and in finding the errors in the final tidal parameters. In particular, the relative error for a harmonic of amplitude A analyzed over a total time span T is approximately $2P(f)/A^2T$, where $P(f)$ is the noise power spectral density at the frequency of that harmonic (Cartwright and Amin, 1986; Munk and Hasselmann, 1964). If excess energy in the tidal bands is not allowed for, the errors can be underestimated by significant amounts.

The main problem with using eqn [53] directly for tidal analysis comes from the fine-scale frequency structure of the tidal forcing, particularly the nodal modulations. Leaving such variations out of eqn [53], and only solving for a few large harmonics, will be inaccurate. But the simplest way of including nodal and other modulations, namely, by including the satellite harmonics in eqn [53], is not possible because the solution will be unstable unless we have 19 years of data. This instability is general and applies whenever we try to solve for the amplitudes of harmonics separated in frequency by less than $1/T$, where T is the record length (Munk and Hasselmann, 1964). This problem is not restricted to the nodal modulation; for example, with only a month of data, we cannot get reliable results for the P_1 and K_1 lines, since they are separated by only 0.15 cycles per month.

All least-squares tidal analysis thus has to include assumptions about tidal harmonics closely spaced in frequency – which comes to an implicit assumption about the smoothness of the admittance. Usually, the admittance is assumed to be constant over frequency ranges of width $1/T$ around the main constituents, summing all harmonics within each such range to form (slowly varying) sinusoidal functions to replace the sines and cosines of eqn [53]. Of course, if we then wish to assign the resulting amplitude to a particular harmonic (say M_2), we need to correct the amplitudes found by the ratio of this sinusoidal function to the single harmonic. All this adds complexity to the existing analysis programs (Foreman, 2004; Pawlowicz et al., 2002; Tamura et al., 1991; Van Camp and Vauterin, 2005; Wenzel, 1996).

A quite different approach to tidal analysis is the *response method*, introduced by Munk and Cartwright (1966). This does not use an expansion of the tidal potential into harmonics, but rather treats it as a time series to be fit to the data, using a set of weights to express the admittance. Lambert (1974) and Merriam (2000) had applied this method to Earth tides, and it has been used for estimating tides from satellite altimetry (Smith et al., 1997).

The basic approach is to find the tides as a weighted sum over the time variations of each spherical harmonic (*not* harmonics in time):

$$y(t)=\sum_{n=2}^{\infty}\sum_{m=-n}^{n}\sum_{l=-L_{nm}}^{L_{nm}} w_{nl}^{m}\left[a_n^m(t-l\Delta)+\mathrm{i}b_n^m(t-l\Delta)\right] \qquad [54]$$

where the $a_n^m(t)$ and $b_n^m(t)$ are the time-varying functions that sum to give the potential in eqn [8]. The complex-valued weights w_{nl}^m are called *response weights*; their Fourier transform gives the admittance $W(f)$. So, for example, a single complex weight for each n and m (i.e., setting $L_{nm}=0$) amounts to assuming a constant W for each degree and order – though even one complex weight can express both amplitude and phase response. Including more weights, with time lags (the sum over l), allows the admittance to vary smoothly with frequency across each tidal band. The lag interval is usually chosen to be 2 days, which makes the admittance smooth over frequencies of greater than 0.5 cpd; note that the weights can include the potential at future times because the admittance is being fit over only a narrow frequency band.

3.06.5.1.1 Predicting tides

All tidal predictions, other than those based on the response method, use a harmonic expansion similar to eqn [11]:

$$x(t)=\sum_{k=1}^{K}A_k\cos\left[2\pi f_k(t-t_0)+\phi_k^0(t_0)+\phi_k\right] \qquad [55]$$

where the A_k's and ϕ_k's are amplitudes and phases (the *harmonic constants*) for whatever is being predicted. The f_k's are the frequencies of the different harmonics, and the ϕ_k^0's are the phases of these at a reference time t_0.

Any user of tidal constants should be aware of two pitfalls relating to the conventions for phase. One is a sign convention: whether positive phases represent lags (time delays), as is true in much of the older literature, or leads. The other is the reference time used. The *local phase* is one choice, in which zero phase (for each harmonic) is at a time at which the potential from that harmonic is locally a maximum. For ocean tides, this phase is usually denoted as κ (with positive phases for lags). For Earth tides, local phase is convenient because on an SNREI Earth, it is zero for gravity, NS tilt, vertical and NS displacement, and areal strain. The other choice is the *Greenwich phase G*, in which the phase is taken to be zero (for each harmonic) at a time at which its potential would be a maximum at 0° longitude. If given for a number of places, this phase provides a 'snapshot' of the distribution of the tides at a particular instant; this phase is the norm in ocean-tide models. Since the time between maximum at Greenwich and maximum at a local place depends only on the

spherical harmonic order m and on the Earth rotating 360° every 24 h, the relationship between κ and G is simple and depends only on the tidal species number m and the longitude ϕ_W; the frequency of the harmonic is not involved. The relationship is conventionally written as

$$G = \kappa - m\phi_W$$

where for both phases, a lag is taken to be positive, and longitude ϕ_W to be positive going west.

The primary complications in predicting the tides come, once again, from the various long-term modulations, notably the nodal modulations discussed in **Section 3.06.2.3**. Classical prediction methods, which used only a few constituents to minimize computation, applied nodal corrections to the A_k's and ϕ_k's of these few harmonics to produce new values that would be valid for (say) each year, a complication, since the corrections themselves change with time.

A more computationally intensive but conceptually simpler approach uses a large number of harmonics in the sum [55], including all satellite harmonics, thus automatically producing the modulations. The amplitudes and phases of a few harmonics are interpolated to give those of all harmonics, again on the assumption that the admittance is smooth, for example, through a spline interpolation of the real and imaginary parts of $W(f)$ (Le Provost et al., 1991).

3.06.6 Earth-Tide Instruments and Measurements

Because the Earth tides are so small, building instruments to detect them has long been a challenge, though digital recording and improvements in transducers have made tidal measurements easier than they used to be.

The earliest measurements were of tidal tilts, and over the years, a wide variety of tiltmeters have been designed; Agnew (1986) described many of them, and the designs have changed little since then. They generally fall into two classes: small instruments that sense the motion of a pendulum or of a bubble in a fluid and larger systems that measure the motion of a fluid in a long pipe. The former are usually referred to as short-base tiltmeters and are now generally installed in boreholes in order to obtain adequate thermal stability and relatively low rates of drift. The latter, called long-base systems, are usually installed in tunnels a few tens of meters long or longer (e.g., d'Oreye and Zürn, 2005), though a very few instruments, several hundred meters in length, have been installed near the ground surface. A similar division exists in strainmeters (Agnew, 1986): there are very short-base systems, many installed in boreholes, longer instruments installed in tunnels, and a few very long instruments, using laser light rather than physical length standards, some at the surface and others underground. One other class of instrument sensitive to tidal deformations is the ring-laser gyroscope (Schreiber et al., 2003), which detects changes in the orientation of the instrument relative to the Earth's spin axis: these involve both tilts and horizontal displacements.

Historically, the second type of Earth tide to be detected was changes in gravity, and many such measurements have been made with a variety of types of gravimeters (Torge, 1989). Most of these used metallic springs, with changes in gravity causing changes in spring length. This mode of operation produced significant phase lags and hysteresis, but such measurements were able to show the widespread effects of load tides (Llubes and Mazzega, 1997) and the effect of the NDFW on tidal data (e.g., Neuberg et al., 1987). Effects caused by spring rheology are much reduced by applying feedback, which was first done in the instrument of LaCoste and Romberg (Harrison and LaCoste, 1978); these instruments, supplied with electronic feedback, remain useful for tidal measurements. The lowest noise and most stable calibration are provided by the superconducting gravimeter (Goodkind, 1999), in which a superconducting sphere is suspended in a magnetic field at liquid-helium temperatures. This gives little drift and very low noise at long periods. At the periods of the semidiurnal tides, the noise on superconducting gravimeters ranges from −130 to −140 dB (relative to $1\,m^2s^{-4}$), with the noise being about 5 dB higher in the diurnal band (Rosat et al., 2004); these levels are about 5 dB below those of spring gravimeters (Cummins et al., 1991). Because the noise is so low, small tidal signals can be measured with great precision; recent examples include the detection of loading from small nonlinear ocean tides (Boy et al., 2004; Merriam, 1995) and the discrimination between the loads from equilibrium and dynamic models for the long-period ocean tides (Boy et al., 2006a; Iwano et al., 2005). The superconducting gravimeter also can provide accurate measurements of the NDFW resonance (Florsch et al., 1994; Sato et al., 1994, 2004; Zürn et al., 1986).

Comparing the size of the load tides in **Figure 8** with the rms body tides in **Figure 6** shows that even the largest loads are but a few percent of the total gravity tide. So, to get accurate measurements of the load tides, the gravimeter must be calibrated with extreme accuracy, at least to a part in 10^3. An even higher level of accuracy is needed if tidal gravity measurements are to discriminate between the predictions of different Earth models. Calibration to this level has proven to be difficult. One method, usable only with spring gravimeters, is to move them to different locations, making gravity measurements over a wide range and interpolating to the small range of the tides. This is not feasible for superconducting gravimeters; they can be calibrated by measuring the response to known moving masses (Achilli et al., 1995), something not generally possible with spring gravimeters because of their complicated mass distribution. A generally applicable method is to place the instrument on a platform that can be moved vertically by known amounts to produce small accelerations of the same size as the tides (Richter et al., 1995). Currently, the commonest calibration method is to operate an absolute gravimeter next to a tidal system for several days and find the scale factor by direct comparison (Francis et al., 1998; Hinderer et al., 1991; Tamura et al., 2005). The calibrations of absolute gravimeters are known to much higher accuracy than is required, so the main source of error is noise in the measurements. Francis and van Dam (2002) confirmed that this method can provide calibrations to 10^{-3}. Baker and Bos (2003) summarized tidal measurements from both spring and superconducting gravimeters; from the greater scatter for the in-phase components of the residual tide, they conclude that systematic errors in the calibration remain, with errors up to 4×10^{-3} in some cases. These authors, and Boy et al. (2003), found too much scatter in

the results to distinguish between different Earth models or to detect any latitude dependence.

The newest procedures for measuring Earth tides are the techniques of space geodesy. Since the induced potential affects satellite orbits, modeling of these provides constraints on k_2; a particularly notable result was that of Ray et al. (2001), who were able to measure the phase lag of the solid-Earth component of the M_2 tide as $0.204^\circ \pm 0.047^\circ$. Positioning techniques such VLBI and GPS are sensitive to all displacements, including tides. VLBI data provided some of the first observations of the body-tide displacements and now have sufficient precision to be sensitive to load tides as well (Haas and Schuh, 1998; Petrov and Ma, 2003; Sovers, 1994); the currently available series can provide the amplitudes of tidal constituents to better than 1 mm. However, VLBI data are available only at a few places, whose number is not increasing.

Continuous GPS data, in contrast, are available from a large, and rapidly increasing, number of locations. The tidal displacements of GPS stations need to be accurately modeled, since any inaccuracy (especially in the vertical) will bias GPS estimates of zenith delay and hence of water vapor (Dach and Dietrich, 2000; Dragert et al., 2000); in addition, for the standard 1-day processing, unmodeled tidal displacements may produce, through aliasing, spurious long-period signals (Penna and Stewart, 2003). All GPS processing software therefore includes models for the body-tide displacements (including the pole tide) and often the load displacements as well.

Tidal motions can be measured with GPS in two ways. One is to process data over short time spans (say, every hour) to produce a time series that can then be analyzed for tidal motions; this is usually called a 'kinematic' method. The other procedure is to include the complex amplitude of some of the larger tidal constituents (in all three directions) as unknowns in the GPS solution, solving for any unmodeled tides. For a single-day solution, the errors will be large because of the close frequency spacing, but keeping the full set of covariances will retain information about what combinations are well constrained. As daily solutions are combined (usually with a Kalman filter estimator), the different harmonics decorrelate, and the tidal estimates converge to a solution.

Baker et al. (1995) and Hatanaka et al. (2001) took the first approach for two local areas, finding good agreement between observed and predicted loads. Khan and Scherneck (2003) also used this method, finding that the best approach was to estimate zenith delays over the same (hourly) time span as the displacement was found for. Schenewerk et al. (2001) used the second method for a global set of stations, taking data from every third day over 3 years; they found generally good agreement with predicted loads except at some high-latitude sites, probably because of using an older ocean-tide model; Allinson et al. (2004) applied this method to tidal loading in Britain. King (2006) compared the two methods for the GPS site at the South Pole (for which only load tides are present), using about 5 years of data. He found that the second method gave better results in the vertical; the two methods had comparable errors in the horizontal, with the precision being somewhat less than 1 mm. The K_1 and K_2 tides give poor results because their frequency is very close to the repeat time of the GPS constellation and its first harmonic. King et al. (2005) used GPS to validate different ocean-tide models around Antarctica, with models specifically designed for this region predicting the loads better than the older global models.

The substantial growth in continuous GPS measurements has meant that measurements of displacement tides are available from many locations, with some regions (Japan, Europe, and the Western United States) having a much higher density of stations than has ever been available for any other type of Earth-tide measurement. As the available data spans have increased, the errors in the tidal estimates have shrunk to a few tenths of a millimeter. Ito et al. (2009) estimated tidal displacements using GPS data from the Japanese GEONET array; Ito and Simons (2011) and Yuan and Chao (2012) had done the same with GPS data from continuous GPS stations in the Western United States. These authors found that, after removing the body and load tides, there are spatially coherent patterns in the residual tidal displacements; **Figure 9** shows ones derived from Yuan and Chao (2012). Since these residuals are not always largest near the coast, it seems unlikely that they are caused by inadequate modeling of the load tides; how to explain them in terms of the laterally heterogeneous models described in **Section 3.06.3.3** remains an open question, as does their relationship to the kind of systematic departures from theoretical tides documented by Langbein (2010) for long-base strain data.

3.06.6.1 Local Distortion of the Tides

Heterogeneities on a much smaller scale than those discussed in **Section 3.06.3.3** can also affect Earth-tide measurements, in ways that make interpretation much more difficult. King and Bilham (1973) and Baker and Lennon (1973) introduced this concept into Earth-tide studies by suggesting that much of the scatter in tidal tilts measurements was caused by strain tides, which were coupled into tilts by the presence of a free surface – for example, the tunnel in which such instruments were often housed. This *cavity effect* (Harrison, 1976) has indeed turned out to be quite important; any strain and tilt measurements not made with surface-mounted long-base instruments require a cavity, which creates a very strong inhomogeneity that produces local deformations, notably rotations and strains that modify the tidal strains and tilts expected for an SNREI Earth. Similar effects are caused by other inhomogeneities such as topography (an irregular free surface), fractures, or changes in rock properties.

We can describe these local effects using coupling matrices (Berger and Beaumont, 1976; King et al., 1976). We denote the actual displacement field in the Earth by $\mathbf{u}$ and divide it into two parts: u_0, which is the displacement that would occur with no cavity (or other inhomogeneity), and the difference $\delta\mathbf{u} = \mathbf{u} - \mathbf{u}_0$. We can perform the same decomposition on the strain tensor $\mathbf{E}$ and the local rotation vector Ω:

$$\mathbf{E} = \mathbf{E}_0 + \delta\mathbf{E} \quad \Omega = \Omega_0 + \delta\Omega$$

The additional deformations $\delta\mathbf{E}$ and $\delta\Omega$ will depend on the details of the irregularity (i.e., the shape of the cavity or the topography) and on $\mathbf{E}_0$, but not on Ω_0 because it is a rigid rotation. If we suppose $\mathbf{E}_0$ to be homogeneous strain, then δE and $\delta\Omega$ at any position are linear functions of $\mathbf{E}_0$ and can be

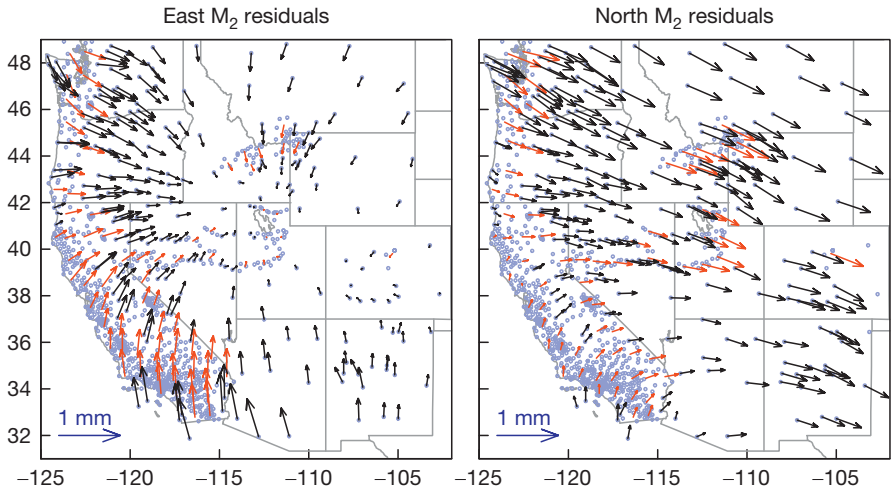


Figure 9 Residual M_2 tides for east and north displacements, represented as phasors relative to the local potential (lags counterclockwise). These come from the GPS results of Yuan and Chao (2012), after subtracting loads from both the GOT4.7 global model and local models for the eastern Pacific, the Gulf of California, Hudson's Bay, and the Arctic Ocean. If there are three or fewer results in a 1° square, they are plotted individually as black arrows; otherwise, a red arrow shows the average value. Blue circles are station locations.

related to it by a fourth-order strain-strain coupling tensor $\mathbf{C_E}$ and a third-order strain-rotation coupling tensor $\mathbf{C_\Omega}$:

$$\delta\mathbf{E} = \mathbf{C_E}\mathbf{E}_0 \quad \delta\Omega = \mathbf{C_\Omega}\mathbf{E}_0 \qquad [56]$$

In general, strain-tilt coupling involves both $\mathbf{C_E}$ and $\mathbf{C_\Omega}$ (Agnew, 1986). Near the surface of the Earth, there are only three independent components to $\mathbf{E}_0$; this fact, and the symmetries of the strain tensors, means that both $\mathbf{C_E}$ and $\mathbf{C_\Omega}$ have nine independent components.

An analytic solution for $\mathbf{C_E}$ and $\mathbf{C_\Omega}$ exists for an ellipsoidal cavity (Eshelby, 1957); Harrison (1976) had described results for the case where one axis of the ellipsoid is vertical: the coupled strains and tilts are largest when measured in the direction along which the cavity is smallest, whether the external strain is in that direction or not. Strain measured along a long narrow cavity is not amplified very much. The limit of this is an infinite horizontal circular tunnel, for which the strains along the tunnel (or shear strains across it) are unaltered, but both vertical and horizontal strains are amplified, as has been observed in cross tunnel measurements of strain tides (Beavan et al., 1979; Itsueli et al., 1975; Takemoto, 1981). Finite-element modeling (Berger and Beaumont, 1976) shows that slight departures from circularity do not much alter the strain from wall to wall, but in a square tunnel, the strains concentrate near the corners; strains along a finite tunnel are undistorted if they are measured more than one tunnel diameter away from an end. Tiltmeters mounted on tunnel walls, or placed on a ledge or near a crack, will be affected by strain-coupled tilts.

Inhomogeneities (including cavities) create distortion but do not add noise. They thus mean that precise comparison between data and models is difficult at best, with the exception of frequency-dependent effects such as the NDFW. There are semianalytic methods for approximating topographic effects (Cayol and Cornet, 1997, 1998; Meertens and Wahr, 1986), and while it is always possible to build a finite-element model of any inhomogeneity (Emter and Zürn, 1985; Sato and Harrison, 1990), not enough detail of the elastic constants is usually known to make results from such a model more than a rough guide. The measurements by Baker (1980), on a pier in the middle of a tunnel, which showed a 5% change in tilts measured 0.5 m apart, imply that such modeling has to be extremely detailed.

One possible way to reduce the importance of coupling is to arrange the geometry to minimize it. Measuring strain along a tunnel is one example; another is measuring tilt in a borehole: since a horizontal strain does not cause rotation of the side of the borehole, attaching the tiltmeter to this should eliminate the cavity effect. An array of such tiltmeters installed to use observed tides to map crustal inhomogeneities (Levine et al., 1989; Meertens et al., 1989) still produced widely scattered results. A subsequent test using closely spaced instruments in nominally uniform geology (Kohl and Levine, 1995) also showed significant differences between boreholes and in one borehole even between two different positions. These results suggest that any short-base measurement of tilt may be significantly affected by local inhomogeneities and so will not be able to be accurately compared with theoretical tidal models.

Sometimes, the aim is to use the tides to determine the effects of the inhomogeneity, so that other signals seen can be corrected. The most notable case is borehole strainmeters, in which the instrument, hole, and grout cause the strain inside the instrument wall to be significantly different from that in the far field. A simplified model that assumes axial symmetry (Gladwin and Hart, 1985) provides two coupling coefficients, one for areal and one for shear strain. This can be checked using the tides (Hart et al., 1996), which also allow estimation of a full coupling tensor, along the lines of equation [56]; Roeloffs (2010) used a version of this in calibrating a later generation of borehole strainmeters.

References

Achilli V, Baldi P, Casula G, et al. (1995) A calibration system for superconducting gravimeters. *Bulletin Géodésique* 69: 73–80.

Agnew DC (1983) Conservation of mass in tidal loading computations. *Geophysical Journal of the Royal Astronomical Society* 72: 321–325.

Agnew DC (1986) Strainmeters and tiltmeters. *Reviews of Geophysics* 24: 579–624.

Agnew DC (1996) SPOTL: Some programs for ocean-tide loading. *SIO Reference Series 96-8*. Scripps Institution of Oceanography.

Agnew DC (2012) SPOTL: Some programs for ocean-tide loading. *SIO Technical Report*. Scripps Institution of Oceanography. http://escholarship.org/uc/item/954322pg.

Allinson CR, Clarke PJ, Edwards SJ, King MA, Baker TF, and Cruddace PR (2004) Stability of direct GPS estimates of ocean tide loading. *Geophysical Research Letters* 31(15): L15603. http://dx.doi.org/10.1029/2004GL020588.

Amoruso A, Botta V, and Crescentini L (2012) Free core resonance parameters from strain data: sensitivity analysis and results from the Gran Sasso (Italy) extensometers. *Geophysical Journal International* 189: 923–936. http://dx.doi.org/10.1111/j.1365-246X.2012.05440.x.

Andersen OB, Woodworth PL, and Flather RA (1995) Intercomparison of recent ocean tide models. *Journal of Geophysical Research* 100: 25,261–25,282.

Baker TF (1980) Tidal tilt at Llanrwst, north Wales: Tidal loading and earth structure. *Geophysical Journal of the Royal Astronomical Society* 62: 269–290.

Baker TF (1984) Tidal deformations of the Earth. *Science Progress Oxford* 69: 197–233.

Baker TF and Bos MS (2003) Validating Earth and ocean tide models using tidal gravity measurements. *Geophysical Journal International* 152: 468–485.

Baker TF, Curtis DJ, and Dodson AH (1995) Ocean tide loading and GPS. *GPS World* 6(5): 54–59.

Baker TF and Lennon GW (1973) Tidal tilt anomalies. *Nature* 243: 75–76.

Bartels J (1957/1985) Tidal forces (English translation). In: Harrison JC (ed.) *Earth Tides*, pp. 25–63. New York: Van Nostrand Reinhold.

Beavan J, Bilham R, Emter D, and King G (1979) Observations of strain enhancement across a fissure. *Veroeff Deutsche Geodaetische Kommission, Reihe B* 231: 47–58.

Benjamin D, Wahr JM, Ray RD, Egbert GD, and Desai SD (2006) Constraints on mantle anelasticity from geodetic observations, and implications for the J2 anomaly. *Geophysical Journal International* 165: 3–16. http://dx.doi.org/10.1111/j.1365-246X.2006.02915.x.

Berger J (1975) A note on thermoelastic strains and tilts. *Journal of Geophysical Research* 80: 274–277.

Berger J and Beaumont C (1976) An analysis of tidal strain observations from the United States of America: II. The inhomogeneous tide. *Bulletin of Seismological Society of America* 66: 1821–1846.

Bindschadler R, Choi H, Wichlacz A, et al. (2011) Getting around Antarctica: New high-resolution mappings of the grounded and freely-floating boundaries of the Antarctic ice sheet created for the International Polar Year. *Cryosphere Discussion* 5: 183–227. http://dx.doi.org/10.5194/tcd-5-183-2011.

Blewitt G (2003) Self-consistency in reference frames, geocenter definition, and surface loading of the solid Earth. *Journal of Geophysical Research* 108: 2103. http://dx.doi.org/10.1029/2002JB002082.

Bos MS and Baker TF (2005) An estimate of the errors in gravity ocean tide loading computations. *Journal of Geodesy* 79: 50–63.

Boy J-P, Llubes M, Hinderer J, and Florsch N (2003) A comparison of tidal ocean loading models using superconducting gravimeter data. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB002050, ETG 6-1–6-17.

Boy J-P, Llubes M, Ray R, Hinderer J, and Florsch N (2006a) Validation of long-period oceanic tidal models with superconducting gravimeters. *Journal of Geodynamics* 41: 112–118.

Boy J-P, Llubes M, Ray R, et al. (2004) Non-linear oceanic tides observed by superconducting gravimeters in Europe. *Journal of Geodynamics* 38: 391–405.

Boy J-P, Ray R, and Hinderer J (2006b) Diurnal atmospheric tide and induced gravity variations. *Journal of Geodynamics* 41: 253–258.

Broucke RA, Zürn W, and Slichter LB (1972) Lunar tidal acceleration on a rigid Earth. *American Geophysical Union, the Geophysical Monograph* 16: 319–324.

Brush SG (1996) *Nebulous Earth: The Origin of the Solar System and the Core of the Earth from Laplace to Jeffreys*. Cambridge: Cambridge University Press.

Brzezínski A and Capitaine N (1993) The use of the precise observations of the Celestial Ephemeris Pole in the analysis of geophysical excitation of earth rotation. *Journal of Geophysical Research* 98: 6667–6675. http://dx.doi.org/10.1029/92JB02874.

Bullesfeld F-J (1985) Ein Beitrag Zur Harmonischen Darstellung Des Gezeitenerzeugenden Potentials. *Deutsche Geodaetische Kommission, Reihe C* 31: 3–103.

Cartwright DE (1977) Oceanic tides. *Reports on Progress in Physics* 40: 665–708.

Cartwright DE (1999) *Tides: A Scientific History*. New York: Cambridge University Press.

Cartwright DE and Amin M (1986) The variances of tidal harmonics. *Deutsche Hydrographische Zeitschrift* 39: 235–253.

Cartwright DE and Edden AC (1973) Corrected tables of tidal harmonics. *Geophysical Journal of the Royal Astronomical Society* 33: 253–264.

Cartwright D, Munk W, and Zetler B (1969) Pelagic tidal measurements: A suggested procedure for analysis. *EOS. Transactions of the American Geophysical Union* 50: 472–477.

Cartwright DE and Tayler RJ (1971) New computations of the tide-generating potential. *Geophysical Journal of the Royal Astronomical Society* 23: 45–74.

Cayol V and Cornet FH (1997) 3D mixed boundary elements for elastic deformation field analysis. *International Journal of Rock Mechanics and Mining Science and Geomechanics Abstracts* 34: 275–287.

Cayol V and Cornet FH (1998) Effect of topography on the interpretation of the deformation field of prominent volcanoes: Application to Etna. *Geophysical Research Letters* 25: 1979–1982.

Chapman S and Lindzen RS (1970) *Atmospheric Tides: Gravitational and Thermal*. New York: Gordon and Breach.

Colosi JA and Munk WH (2006) Tales of the venerable Honolulu tide gauge. *Journal of Physical Oceanography* 36: 967–996.

Cummins P, Wahr JM, Agnew DC, and Tamura Y (1991) Constraining core undertones using stacked IDA gravity records. *Geophysical Journal International* 106: 189–198.

Dach R and Dietrich R (2000) Influence of the ocean loading effect on GPS derived precipitable water vapor. *Geophysical Research Letters* 27: 2953–2956. http://dx.doi.org/10.1029/1999GL010970.

Dahlen FA (1993) Effect of the Earth's ellipticity on the lunar potential. *Geophysical Journal International* 113: 250–251.

Dahlen FA and Tromp J (1998) *Theoretical Global Seismology*. Princeton, NJ: Princeton University Press.

Dehant V, Defraigne P, and Wahr JM (1999) Tides for a convective Earth. *Journal of Geophysical Research* 104: 1035–1058.

Desai SD (2002) Observing the pole tide with satellite altimetry. *Journal of Geophysical Research* 107(C11): 3186. http://dx.doi.org/10.1029/2001JC001224.

Desai SD, Wahr JM, and Chao Y (1997) Error analysis of empirical ocean tide models estimated from Topex/Poseidon altimetry. *Journal of Geophysical Research* 102: 25,157–25,172.

Doodson AT (1921) The harmonic development of the tide generating potential. *Proceedings of the Royal Society of London, Series A* 100: 305–329.

d'Oreye NF and Zürn W (2005) Very high resolution long-baseline water-tube tiltmeter to record small signals from Earth free oscillations up to secular tilts. *Review of Scientific Instruments* 76: 024501.

Dragert H, James TS, and Lambert A (2000) Ocean loading corrections for continuous GPS: A case study at the Canadian coastal site Holberg. *Geophysical Research Letters* 27: 2045–2048.

Egbert GD and Erofeeva SY (2002) Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology* 19: 183–204.

Emter D and Zürn W (1985) Observations of local elastic effects on earth tide tilts and strains. In: Harrison JC (ed.) *Earth Tides*, pp. 309–327. New York: Van Nostrand Reinhold.

Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion and related problems. *Proceedings of the Royal Society of London, Series A* 241: 376–396.

Farrell WE (1972) Deformation of the earth by surface loads. *Reviews of Geophysics* 10: 761–797.

Farrell WE (1973) Earth tides, ocean tides, and tidal loading. *Philosophical Transactions of the Royal Society A* 272: 253–259.

Florsch N, Chambat F, Hinderer J, and Legros H (1994) A simple method to retrieve the complex eigenfrequency of the Earth's nearly diurnal free wobble: Application to the Strasbourg superconducting gravimeter data. *Geophysical Journal International* 116: 53–63.

Foreman MGG (2004) Manual for tidal heights analysis and prediction. *Pacific Marine Science Report 77-10*. Institute of Ocean Sciences, Patricia Bay, Sidney, B.C.

Francis O and Dehant V (1987) Recomputation of the Green's functions for tidal loading estimations. *Bulletin d'Information des Marées Terrestres* 100: 6962–6986.

Francis O, Niebauer TM, Sasagawa G, Klopping F, and Gschwind J (1998) Calibration of a superconducting gravimeter by comparison with an absolute gravimeter FG5 in Boulder. *Geophysical Research Letters* 25: 1075–1078.

Francis O and van Dam T (2002) Evaluation of the precision of using absolute gravimeters to calibrate superconducting gravimeters. *Metrologia* 39: 485–488.

Fu Y, Freymueller J, and vanDam T (2012) The effect of using inconsistent ocean tidal loading models on GPS coordinate solutions. *Journal of Geodesy* 86: 409–421. http://dx.doi.org/10.1007/s00190-011-0528-1.

Fu GY and Sun WK (2007) Effects of lateral inhomogeneity in a spherical Earth on gravity Earth tides. *Journal of Geophysical Research* 112. http://dx.doi.org/10.1029/2006JB004512.

Garrett C and Munk WH (1971) The age of the tide and the Q of the oceans. *Deep-Sea Research* 18: 493–503.

Gladwin MT and Hart R (1985) Design parameters for borehole strain instrumentation. *Pure and Applied Geophysics* 123: 59–80.

Goodkind JM (1999) The superconducting gravimeter. *Review of Scientific Instruments* 70: 4131–4152.

Gross R (1992) Correspondence between theory and observations of polar motion. *Geophysical Journal International* 109: 162–170.

Guo J, Li Y, Huang Y, Deng H, Xu S, and Ning J (2004) Green's function of the deformation of the Earth as a result of atmospheric loading. *Geophysical Journal International* 159: 53–68. http://dx.doi.org/10.1111/j.1365-246X.2004.02410.x.

Haas R and Schuh H (1998) Ocean loading observed by geodetic VLBI. In: Ducarme B and Paquet P (eds.) *Proceedings of the 13th International Symposium on Earth Tides*, pp. 111–120. Brussels: Observatoire Royal de Belgique.

Harrison JC (1971) New programs for the computation of Earth tides. *Internal Technical Report*. CIRES, University of Colorado.

Harrison JC (1976) Cavity and topographic effects in tilt and strain measurement. *Journal of Geophysical Research* 81: 319–328.

Harrison JC (1985) *Earth Tides*. New York: Van Nostrand Reinhold.

Harrison JC and Herbst K (1977) Thermoelastic strains and tilts revisited. *Geophysical Research Letters* 4: 535–537.

Harrison JC and LaCoste LJB (1978) The measurement of surface gravity. In: Mueller II (ed.) *Applications of Geodesy to Geodynamics: Ninth GEOP Conference*, pp. 239–243. Columbus, OH: Ohio State University, Department of Geodetic Science, rep. 280.

Hart RHG, Gladwin MT, Gwyther RL, Agnew DC, and Wyatt FK (1996) Tidal calibration of borehole strainmeters: Removing the effects of local inhomogeneity. *Journal of Geophysical Research* 101: 25553–25571.

Hartmann T and Wenzel H-G (1995) The HW95 tidal potential catalogue. *Geophysical Research Letters* 22: 3553–3556.

Hatanaka Y, Sengoku A, Sato T, Johnson JM, Rocken C, and Meertens C (2001) Detection of tidal loading signals from GPS permanent array of GSI Japan. *Journal of the Geodetic Society of Japan* 47: 187–192.

Haubrich RA and Munk WH (1959) The pole tide. *Journal of Geophysical Research* 64: 2373–2388.

Herring TA, Mathews PM, and Buffett BA (2002) Modeling of nutation-precession: Very long baseline interferometry results. *Journal of Geophysical Research* 107(B4): 2069. http://dx.doi.org/10.1029/2001JB000165.

Hewitt E and Hewitt R (1979) The Gibbs–Wilbraham phenomenon: An episode in Fourier analysis. *Archive for History of Exact Sciences* 21: 129–169.

Hinderer J, Florsch N, Maekinen J, Legros H, and Faller JE (1991) On the calibration of a superconducting gravimeter using absolute gravity measurements. *Geophysical Journal International* 106: 491–497.

Ito T, Okubo M, and Sagiya T (2009) High resolution mapping of Earth tide response based on GPS data in Japan. *Journal of Geodynamics* 48: 253–259. http://dx.doi.org/10.1016/j.jog.2009.09.012.

Ito T and Simons M (2011) Probing asthenospheric density, temperature, and elastic moduli below the western United States. *Science* 332: 947–951. http://dx.doi.org/10.1126/science.1202584.

Itsueli UJ, Bilham R, Goulty NR, and King GCP (1975) Tidal strain enhancement observed across a tunnel. *Geophysical Journal of the Royal Astronomical Society* 42: 555–564.

Iwano S, Fukuda Y, Sato T, Tamura Y, Matsumoto K, and Shibuya K (2005) Long-period tidal factors at Antarctica Syowa Station determined from 10 years of superconducting gravimeter data. *Journal of Geophysical Research* 110: B10403. http://dx.doi.org/10.1029/2004JB003551.

Jeffreys H (1976) *The Earth: Its Origin, History and Physical Constitution*. Cambridge: Cambridge University Press.

Jeffreys H and Vincente RO (1957) The theory of nutation and the variation of latitude. *Monthly Notices of the Royal Astronomical Society* 117: 142–161.

Jentzsch G (1997) Earth tides and Ocean tidal loading. In: Wilhelm H, Zürn W, and Wenzel HG (eds.) *Tidal Phenomena*, pp. 145–171. Berlin: Springer-Verlag.

Jentzsch G, Jahr T, and Kroner C (2009) Special issue on "New challenges in Earth's dynamics", Proceedings of the 16th International Symposium on Earth Tides-Joined Meeting of Sub-commissions 3.1 on Earth Rotation and Earth Tides, 3.2 on Crustal Deformation, 3.3 on Geophysical fluids, and the Global Geodynamics Project (GGP), held in Jena, Germany, 1–5 September 2008, *Journal of Geodynamics* 48: 107–109. http://dx.doi.org/10.1016/j.jog.2009.09.041.

Kamigaichi O (1998) Green functions for the earth at borehole installation sensor depths for surface point load. *Papers in Meteorology and Geophysics* 48: 89–100.

Khan SA and Scherneck H-G (2003) The M_2 ocean tide loading wave in Alaska: Vertical and horizontal displacements, modelled and observed. *Journal of Geodesy* 77: 117–127. http://dx.doi.org/10.1007/s00190-003-0312-y.

King M (2006) Kinematic and static GPS techniques for estimating tidal displacements with application to Antarctica. *Journal of Geodynamics* 41: 77–86.

King GCP and Bilham R (1973) Tidal tilt measurement in Europe. *Nature* 243: 74–75.

King MA and Padman L (2005) Accuracy assessment of ocean tide models around Antarctica. *Geophysical Research Letters* 32: L23608. http://dx.doi.org/10.1029/2005GL023901.

King MA, Penna NT, Clarke PJ, and King EC (2005) Validation of ocean tide models around Antarctica using onshore GPS and gravity data. *Journal of Geophysical Research* 110: B08401. http://dx.doi.org/10.1029/2004JB003390.

King GCP, Zürn W, Evans R, and Emter D (1976) Site correction for long-period seismometers, tiltmeters, and strainmeters. *Geophysical Journal of the Royal Astronomical Society* 44: 405–411.

Kohl ML and Levine J (1995) Measurement and interpretation of tidal tilts in a small array. *Journal of Geophysical Research* 100: 3929–3941.

Ku LF, Greenberg DA, Garrett C, and Dobson FW (1985) The nodal modulation of the M_2 tide in the Bay of Fundy and Gulf of Maine. *Science* 230: 69–71.

Kudryavtsev SM (2004) Improved harmonic development of the Earth tide-generating potential. *Journal of Geodesy* 77: 829–838.

Lambert A (1974) Earth tide analysis and prediction by the response method. *Journal of Geophysical Research* 79: 4952–4960.

Langbein J (2010) The effect of error in theoretical earth tide on calibration of borehole strainmeters. *Geophysical Research Letters* 37: L21303. http://dx.doi.org/10.1029/2010GL044454.

Latychev K, Mitrovica J, Ishii M, Chan N, and Davis J (2009) Body tides on a 3-D elastic earth: Toward a tidal tomography. *Earth and Planetary Science Letters* 277: 86–90. http://dx.doi.org/10.1016/j.epsl.2008.10.008.

Le Provost C, Bennett AF, and Cartwright DE (1995) Ocean tides for and from Topex/Poseidon. *Science* 267: 639–642.

Le Provost C, Lyard F, and Molines J (1991) Improving ocean tide predictions by using additional semidiurnal constituents from spline interpolation in the frequency domain. *Geophysical Research Letters* 18: 845–848.

Levine J, Meertens C, and Busby R (1989) Tilt observations using borehole tiltmeters: 1. Analysis of tidal and secular tilt. *Journal of Geophysical Research* 94: 574–586.

Llubes M and Mazzega P (1997) Testing recent global ocean tide models with loading gravimetric data. *Progress in Oceanography* 40: 369–383.

Longman IM (1959) Formulas for computing the tidal accelerations due to the Moon and the Sun. *Journal of Geophysical Research* 64: 2351–2355.

Love AEH (1911) *Some Problems of Geodynamics*. Cambridge: Cambridge University Press.

Lyard F, Lefevre F, Letellier T, and Francis O (2006) Modelling the global ocean tides: Modern insights from FES2004. *Ocean Dynamics* 56: 394–415.

Mathews PM, Buffett BA, and Shapiro II (1995a) Love numbers for diurnal tides: Relation to wobble admittances and resonance expansion. *Journal of Geophysical Research* 100: 9935–9948.

Mathews PM, Buffett BA, and Shapiro II (1995b) Love numbers for a rotating spheroidal Earth: New definitions and numerical values. *Geophysical Research Letters* 22: 579–582. http://dx.doi.org/10.1029/95GL00161.

Mathews PM, Dehant V, and Gipson JM (1997) Tidal station displacements. *Journal of Geophysical Research* 102: 20,469–20,477.

Mathews PM and Guo JY (2005) Viscoelectromagnetic coupling in precession-nutation theory. *Journal of Geophysical Research* 110(B2): B02402. http://dx.doi.org/10.1029/2003JB002915.

Mathews PM, Herring TA, and Buffett BA (2002) Modeling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior. *Journal of Geophysical Research* 107(B4): 2068. http://dx.doi.org/10.1029/2001JB000390.

Matsumoto K, Sato T, Takanezawa T, and Ooe M (2001) GOTIC2: A program for computation of oceanic tidal loading effect. *Journal of the Geodetic Society of Japan* 47: 243–248.

Matsumoto K, Takanezawa T, and Ooe M (2000) Ocean tide models developed by assimilating Topex/Poseidon altimeter data into hydrodynamical model: A global model and a regional model around Japan. *Journal of Oceanography* 56: 567–581.

Meertens C, Levine J, and Busby R (1989) Tilt observations using borehole tiltmeters: 2. Analysis of data from Yellowstone National Park. *Journal of Geophysical Research* 94: 587–602.

Meertens CM and Wahr JM (1986) Topographic effect on tilt, strain, and displacement measurements. *Journal of Geophysical Research* 91: 14,057–14,062.

Melchior P (1983) *The Tides of the Planet Earth*. New York: Pergamon.

Merriam JB (1980) The series computation of the gravitational perturbation due to an ocean tide. *Physics of the Earth and Planetary Interiors* 23: 81–86.

Merriam JB (1985) Toroidal Love numbers and transverse stress at the Earth's surface. *Journal of Geophysical Research* 90: 7795–7802.

Merriam JB (1986) Transverse stress Green's functions. *Journal of Geophysical Research* 91: 13903–13913.

Merriam JB (1992) An ephemeris for gravity tide predictions at the nanogal level. *Geophysical Journal International* 108: 415–422.
Merriam JB (1995) Nonlinear tides observed with the superconducting gravimeter. *Geophysical Journal International* 123: 529–540.
Merriam JB (2000) The response method applied to the analysis of superconducting gravimeter data. *Physics of the Earth and Planetary Interiors* 121: 289–299.
Metivier L, Greff-Lefftz M, and Diament M (2005) A new approach to computing accurate gravity time variations for a realistic earth model with lateral heterogeneities. *Geophysical Journal International* 162: 570–574. http://dx.doi.org/10.1111/j.1365-246X.2005.02692.x.
Metivier L, Greff-Lefftz M, and Diament M (2006) Mantle lateral variations and elastogravitational deformations – I. Numerical modelling. *Geophysical Journal International* 167: 1060–1076. http://dx.doi.org/10.1111/j.1365-246X.2006.03159.x.
Miller SP and Wunsch C (1973) The pole tide. *Nature Physical Science* 246: 98–102.
Mitrovica JX, Davis JL, and Shapiro II (1994) A spectral formalism for computing three-dimensional deformations due to surface loads: 1. Theory. *Journal of Geophysical Research* 99: 7057–7074.
Molodenskii MS (1961) The theory of nutations and diurnal earth tides. *Belgian Royal Observatory Communications Series Geophysics Series* 58: 25–56.
Molodensky SM (1977) The influence of horizontal inhomogeneities in the mantle on the amplitude of tidal oscillations. *Izvestiya Physics of the Solid Earth* 13: 77–80.
Müller G (1977) Thermoelastic deformations of a half-space: A Green's function approach. *Journal of Geophysics* 43: 761–769.
Munk WH and Cartwright DE (1966) Tidal spectroscopy and prediction. *Philosophical Transactions of the Royal Society A* 259: 533–581.
Munk WH and Hasselmann K (1964) Super-resolution of tides. In: *Studies on Oceanography; A Collection of Papers Dedicated to Koji Hidaka in Commemoration of His Sixtieth Birthday*, pp. 339–344. Tokyo: University of Tokyo Press.
Munk WH, Zetler BD, and Groves GW (1965) Tidal cusps. *Geophysical Journal* 10: 211–219.
Neuberg J, Hinderer J, and Zürn W (1987) Stacking gravity tide observations in central Europe for the retrieval of the complex eigenfrequency of the nearly diurnal free wobble. *Geophysical Journal of the Royal Astronomical Society* 91: 853–868.
Padman L and Erofeeva S (2004) A barotropic inverse tidal model for the Arctic Ocean. *Geophysical Research Letters* 31: L02303. http://dx.doi.org/10.1029/2003GL01900.
Padman L, Fricker HA, Coleman R, Howard S, and Erofeeva L (2002) A new tide model for the Antarctic ice shelves and seas. *Annals of Glaciology* 34: 247–254.
Pagiatakis SD (1990) The response of a realistic Earth to ocean tide loading. *Geophysical Journal International* 103: 541–560.
Pawlowicz R, Beardsley B, and Lentz S (2002) Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers and Geosciences* 28: 929–937.
Penna N, Bos M, Baker T, and Scherneck HG (2008) Assessing the accuracy of predicted ocean tide loading displacement values. *Journal of Geodesy* 82: 893–907. http://dx.doi.org/10.1007/s00190-008-0220-2.
Penna NT and Stewart MP (2003) Aliased tidal signatures in continuous GPS height time series. *Geophysical Research Letters* 30. http://dx.doi.org/10.1029/2003GL018828, SDE 1–1.
Pétit G and Luzum B (2010), IERS Conventions (2010), *IERS Technical Note 36*. Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main.
Petrov L and Ma C (2003) Study of harmonic site position variations determined by very long baseline interferometry. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB001801, ETG 5–1.
Ponchaut F, Lyard F, and LeProvost C (2001) An analysis of the tidal signal in the WOCE sea level dataset. *Journal of Atmospheric and Oceanic Technology* 18: 77–91.
Ray RD (1998) Ocean self-attraction and loading in numerical tidal models. *Marine Geodesy* 21: 181–192.
Ray RD (2001) Resonant third-degree diurnal tides in the seas off Western Europe. *Journal of Physical Oceanography* 31: 3581–3586.
Ray RD, Bills BG, and Chao BF (1999) Lunar and solar torques on the oceanic tides. *Journal of Geophysical Research* 104: 17653–17659.
Ray RD, Eanes RJ, and Lemoine FG (2001) Constraints on energy dissipation in the Earth's body tide from satellite tracking and altimetry. *Geophysical Journal International* 144: 471–480.
Ray RD and Egbert GD (2004) The global S_1 tide. *Journal of Physical Oceanography* 34: 1922–1935.
Ray RD and Egbert GD (2012) Fortnightly Earth rotation, ocean tides and mantle anelasticity. *Geophysical Journal International* 189: 400–413. http://dx.doi.org/10.1111/j.1365-246X.2012.05351.x.
Ray RD, Egbert GD, and Erofeeva SY (2011) Tide predictions in shelf and coastal waters: status and prospects. In: Vignudelli S, Kostianoy AG, Cipollini P, and Benveniste J (eds.) *Coastal Altimetry*, pp. 191–216. New York: Springer-Verlag.
Ray RD and Ponte RM (2003) Barometric tides from ECMWF operational analyses. *Annales Geophysicae* 21: 1897–1910.
Ray RD and Poulose S (2005) Terdiurnal surface-pressure oscillations over the continental United States. *Monthly Weather Review* 133: 2526–2534.
Ray RD and Sanchez BV (1989) Radial deformation of the earth by oceanic tidal loading. *NASA Tech. Mem. TM-100743*, Goddard Space Flight Center, Greenbelt, MD.
Richter B, Wilmes H, and Nowak I (1995) The Frankfurt calibration system for relative gravimeters. *Metrologia* 32: 217–223.
Rignot E, Mouginot J, and Scheuchl B (2011) Antarctic grounding line mapping from differential satellite radar interferometry. *Geophysical Research Letters* 38: L10504. http://dx.doi.org/10.1029/2011GL047109.
Roeloffs E (2010) Tidal calibration of PBO borehole strainmeters: The roles of vertical and shear coupling. *Journal of Geophysical Research* 115: B06405. http://dx.doi.org/10.1029/2009JB006407.
Roosbeek F (1995) RATGP95: A harmonic development of the tide-generating potential using an analytical method. *Geophysical Journal International* 126: 197–204.
Rosat S, Hinderer J, Crossley D, and Boy JP (2004) Performance of superconducting gravimeters from long-period seismology to tides. *Journal of Geodynamics* 38: 461–476.
Rosat S and Lambert SB (2009) Free core nutation resonance parameters from VLBI and superconducting gravimeter data. *Astronomy and Astrophysics* 503: 287–291. http://dx.doi.org/10.1051/0004-6361/200811489.
Sato T and Harrison JC (1990) Local effects on tidal strain measurements at Esashi, Japan. *Geophysical Journal International* 102: 513–526.
Sato T, Tamura Y, Higashi T, et al. (1994) Resonance parameters of the free core nutation measured from three superconducting gravimeters in Japan. *Journal of Geomagnetism and Geoelectricity* 46: 571–586.
Sato T, Tamura Y, Matsumoto K, Imanishi Y, and McQueen H (2004) Parameters of the fluid core resonance inferred from superconducting gravimeter data. *Journal of Geodynamics* 38: 375–389. http://dx.doi.org/10.1016/j.jog.2004.07.016.
Schenewerk MS, Marshall J, and Dillinger W (2001) Vertical ocean-loading deformations derived from a global GPS network. *Journal of the Geodetic Society of Japan* 47: 237–242.
Schreiber KU, Klugel T, and Stedman GE (2003) Earth tide and tilt detection by a ring laser gyroscope. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2001JB000569.
Shum CK, Woodworth PL, Andersen OB, et al. (1997) Accuracy assessment of recent ocean tide models. *Journal of Geophysical Research* 102: 25,173–25,194.
Smith A, Ambrosius B, Wakker K, Woodworth P, and Vassie J (1997) Comparison between the harmonic and response methods of tidal analysis using Topex/Poseidon altimetry. *Journal of Geodesy* 71: 695–703. http://dx.doi.org/10.1007/s001900050137.
Smith ML and Dahlen FA (1981) The period and Q of the Chandler wobble. *Geophysical Journal of the Royal Astronomical Society* 64: 223–281.
Sovers OJ (1994) Vertical ocean loading amplitudes from VLBI measurements. *Geophysical Research Letters* 21: 357–360.
Standish EM, Newhall XX, Williams JG, and Yeomans DK (1992) Orbital ephemerides of the Sun, Moon and planets. In: Seidelmann PK (ed.) *Explanatory Supplement to the Astronomical Almanac*, pp. 279–323. Sausalito, CA: University Science Books.
Steinberger B and Holme R (2008) Mantle flow models with core–mantle boundary constraints and chemical heterogeneities in the lowermost mantle. *Journal of Geophysical Research* 113: B05403. http://dx.doi.org/10.1029/2007JB005080.
Takemoto S (1981) Effects of local inhomogeneities on tidal strain measurements. *Bulletin of the Disaster Prevention Research Institute, Kyoto University* 31: 211–237.
Tamura Y (1982) A computer program for calculating the tide generating force. *Publications of the International Latitude Observatory, Mizusawa* 16: 1–20.
Tamura Y (1987) A harmonic development of the tide-generating potential. *Bulletin d'Information des Marées Terrestres* 99: 6813–6855.
Tamura Y, Sato T, Fukuda Y, and Higashi T (2005) Scale factor calibration of a superconducting gravimeter at Esashi Station, Japan, using absolute gravity measurements. *Journal of Geodesy* 78: 481–488.
Tamura Y, Sato T, Ooe M, and Ishiguro M (1991) A procedure for tidal analysis with a Bayesian information criterion. *Geophysical Journal International* 104: 507–516.
Torge W (1989) *Gravimetry*. Berlin: Walter de Gruyter Verlag.
Trupin A and Wahr J (1990) Spectroscopic analysis of global tide gauge sea level data. *Geophysical Journal International* 100: 441–453.
Van Camp M and Vauterin P (2005) Tsoft: Graphical and interactive software for the analysis of time series and Earth tides. *Computers and Geosciences* 31: 631–640.

Wahr JM (1981a) Body tides on an elliptical, rotating, elastic and oceanless Earth. *Geophysical Journal of the Royal Astronomical Society* 64: 677–703.
Wahr JM (1981b) A normal mode expansion for the forced response of a rotating Earth. *Geophysical Journal of the Royal Astronomical Society* 64: 651–675.
Wahr JM and Bergen Z (1986) The effects of mantle and anelasticity on nutations, earth tides, and tidal variations in rotation rate. *Geophysical Journal* 87: 633–668.
Wahr JM and Sasao T (1981) A diurnal resonance in the ocean tide and in the Earth's load response due to the resonant free "core nutation" *Geophysical Journal of the Royal Astronomical Society* 64: 747–765.
Wang R (1997) Tidal response of the solid Earth. In: Wilhelm H, Zürn W, and Wenzel HG (eds.) *Tidal Phenomena*, pp. 27–57. Berlin: Springer-Verlag.
Warburton RJ and Goodkind JM (1976) Search for evidence of a preferred reference frame. *Astrophysical Journal* 208: 881–886.
Wenzel H-G (1996) The nanogal software: Earth tide data processing package ETERNA 3.3. *Bulletin d'Information des Marées Terrestres* 124: 9425–9439.
Wessel P and Smith WHF (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research* 101: 8741–8743.
Wilhelm H (1983) Earth's flattening effect on the tidal forcing field. *Journal of Geophysics* 52: 131–135.
Wilhelm H (1986) Spheroidal and torsional stress coefficients. *Journal of Geophysics* 55: 423–432.
Wilhelm H, Zürn W, and Wenzel HG (1997) *Tidal Phenomena*. Berlin: Springer-Verlag.
Xi Q (1987) A new complete development of the tide-generating potential for the epoch J2000.0. *Bulletin d'Information des Marées Terrestres* 99: 6766–6812.
Yuan L and Chao B (2012) Analysis of tidal signals in surface displacement measured by a dense continuous GPS array. *Earth and Planetary Science Letters* 355–356: 255–261. http://dx.doi.org/10.1016/j.epsl.2012.08.035.
Zürn W (1997a) Earth tide observations and interpretation. In: Wilhelm H, Zürn W, and Wenzel HG (eds.) *Tidal Phenomena*, pp. 77–94. Berlin: Springer-Verlag.
Zürn W (1997b) The nearly-diurnal free wobble-resonance. In: Wilhelm H, Zürn W, and Wenzel HG (eds.) *Tidal Phenomena*, pp. 95–109. Berlin: Springer-Verlag.
Zürn W, Rydelek PA, and Richter B (1986) The core-resonance effect in the record from the superconducting gravimeter at Bad Homburg. In: Viera R (ed.) *Proceedings of the 10th International Symposium on Earth Tides*, pp. 141–147. Madrid: Consejo Superior de Investigaciones Cientificas.

3.07 Glacial Isostatic Adjustment and the Long-Wavelength Gravity Field

ME Tamisiea, National Oceanography Centre, Liverpool, UK
JX Mitrovica, Harvard University, Cambridge, MA, USA
K Latychev, University of Toronto, Toronto, Ontario, Canada

© 2015 Elsevier B.V. All rights reserved.

3.07.1	Introduction	179
3.07.2	$\dot{J}_2$: A Review of Early GIA Research	179
3.07.3	Assessment of the Long-Wavelength Gravity Trends	181
3.07.4	Degree-Two Trends in Gravity Revisited	182
3.07.4.1	Uncertainties in the $\dot{J}_2$ Constraint	183
3.07.4.2	Geoid Height Change: The Influence of Rotational Feedback	183
3.07.5	The Influence of Lateral Variations in Mantle Viscosity	187
3.07.6	Summary	188
References		189

3.07.1 Introduction

The adjustment of the Earth in response to the glacial cycles of the Late Pleistocene ice age is reflected in a wide range of geophysical observations, including sea-level variations and anomalies in the gravitational field and rotational state of the planet. Early analyses of glacial isostatic adjustment (GIA), dating up to the early 1980s, involved geologic records of Holocene sea-level trends (Cathles, 1975; McConnell, 1968; Peltier and Andrews, 1976; Wu and Peltier, 1983), regional maps of static gravity anomalies obtained by land-based surveys in previously glaciated regions such as Laurentia and Fennoscandia (Cathles, 1975; Walcott, 1972; Wu and Peltier, 1983), and astronomical measurements of secular trends in the magnitude and orientation of the rotation vector (Nakiboglu and Lambeck, 1980; Sabadini and Peltier, 1981; Wu and Peltier, 1984).

As space-geodetic techniques became increasingly common, these observations placed greater constraints on the GIA process. These included early estimates of 3-D crustal velocities using very long-baseline interferometry data (Argus, 1996; James and Lambert, 1993; Mitrovica et al., 1993, 1994; Peltier, 1995). The spatial coverage of measurements of 3-D crustal motion greatly improved with the advent and improvement of surveying using the global positioning system (GPS), and this coverage became an important component of regional studies of GIA (e.g., Calais et al., 2006; Dietrich et al., 2010; Johansson et al., 2002; Milne et al., 2001; Sella et al., 2007; Thomas et al., 2011). One of the earliest targets for space-geodetic missions was the time variation of the long-wavelength components of the Earth's gravitational field. These components, the so-called Stokes coefficients, were the focus of several early GIA modeling studies (e.g., Ivins et al., 1993; Mitrovica and Peltier, 1993; Sabadini et al., 2002; Yoder et al., 1983).

This chapter focuses on the modeling of the ice age contribution to these longest-wavelength components of the Earth's geopotential. We begin with a discussion of the 'early' GIA studies of this kind, with a particular emphasis on the evolving goals of these analyses. We then move on to advances in such analysis, including the incorporation of rotational effects and 3-D variations in viscosity into predictions that have traditionally assumed nonrotating, spherically symmetrical viscoelastic Earth models. Since the launch of the Gravity Recovery and Climate Experiment (GRACE) satellite mission (Tapley et al., 2004), much of the focus has shifted to the new regional estimates of the time-variable geopotential (Wahr, **Chapter 3.08**). However, it is still appropriate for such a review given that these long-wavelength contributions from GIA can introduce a background bias in global averages obtained from these geodetic observations if they are not properly considered.

3.07.2 $\dot{J}_2$: A Review of Early GIA Research

We begin by reproducing the usual spherical harmonic decomposition of the secular trend in the geoid height anomaly, $\dot{G}$:

$$\dot{G}(\theta, \phi, t_p) = a\sum_{l=0}^{\infty}\sum_{m=0}^{l}\widetilde{P}_{lm}(\cos\theta)\left[\dot{C}_{lm}(t_p)\cos m\phi + \dot{S}_{lm}(t_p)\sin m\phi\right] \quad [1]$$

where θ and ϕ are the colatitude and east longitude, respectively, t_p is the time at present day, a is the Earth's radius, and $\widetilde{P}_{lm}$ is the normalized Legendre polynomial of degree l and order m. The constants $\dot{C}_{lm}$ and $\dot{S}_{lm}$ are the Stokes coefficients. The zonal harmonics, $\dot{J}_l$, are related to the coefficients $\dot{C}_{l,m=0}$ by

$$\dot{J}_l(t_p) = -\sqrt{2l+1}\,\dot{C}_{l,m=0}(t_p) \quad [2]$$

We note that the $\dot{J}_2$ datum is linearly proportional to the rate of change of the rotation rate, $\dot{\Omega}$ ($\dot{J}_2 \sim -0.5\dot{\Omega}/\Omega$; Wu and Peltier, 1984).

Since the pioneering study of Yoder et al. (1983), who estimated the $\dot{J}_2$ coefficient using LAGEOS satellite data, the geophysical goals of GIA analyses associated with these coefficients have changed. The earliest studies assumed that GIA was the sole contributor to the $\dot{J}_2$ signal, and the observation was used to estimate the bulk lower mantle viscosity of the Earth

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00064-6

(e.g., Alexander, 1983; Peltier, 1983, 1985; Rubincam, 1984; Wu and Peltier, 1984; Yoder et al., 1983; Yuen and Sabadini, 1985). This application is illustrated in **Figure 1(a)**, where we compare numerical predictions of $\dot{J}_2$ (or alternatively, $\dot{\Omega}/\Omega$) due to ongoing GIA with the observational constraint cited by Nerem and Klosko (1996). (The latter is consistent with the earlier Yoder et al. estimate of -3×10^{-11}year^{-1}.)

The predictions are based on standard GIA methods for spherically symmetrical (i.e., depth-varying), self-gravitating, Maxwell viscoelastic Earth models (Mitrovica and Peltier, 1989), and they include a global ice history modified from the ICE-3G deglaciation geometry (Tushingham and Peltier, 1991) as well as a complementary water load. The calculations assume a PREM elastic structure (Dziewonski and Anderson, 1981), lithospheric thickness of 100 km, and isoviscous upper and lower mantle regions (henceforth ν_{UM} and ν_{LM}, respectively, with the boundary between the two taken to be 670 km depth); the former is fixed to a value of 5×10^{20} Pa s, while the latter is a free parameter of the modeling that is varied over two orders of magnitude (as given by the abscissa scale). The solid line on the figure has the classic 'inverted parabola' form common to many predictions of ongoing GIA trends as a function of ν_{LM}. In the case of a weak lower mantle viscosity ($\nu_{\mathrm{LM}} < 4 \times 10^{21}$ Pa s), the planetary model would have relaxed close to equilibrium since the end of the ICE-3G deglaciation (5 ky BP), and thus, the predicted present-day rates are small. Models with a stiff lower mantle ($\nu_{\mathrm{LM}} > 4 \times 10^{22}$ Pa s) are characterized by a slow rate of adjustment throughout the glacial cycle. Hence, peak rates are predicted for models intermediate to these extremes.

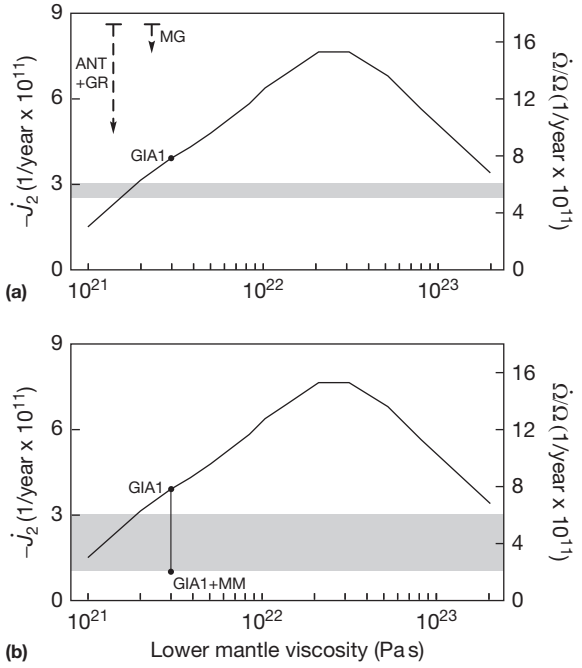


Figure 1 (a) Predictions of $\dot{J}_2$ due to GIA as a function of the lower mantle viscosity of the Earth model (abscissa). The right ordinate scale refers to the normalized rate of change of the rotation rate, $\dot{\Omega}/\Omega$. The shaded region represents the observational constraint on $\dot{J}_2$ derived from satellite data by Nerem and Klosko (1996). The arrows at the top left indicate the $\dot{J}_2$ signal associated with a net twentieth-century melting from the Antarctic and Greenland ice complexes of 1 mm year^{-1} global mean sea-level rise (ANT+GR) and the melting of small ice sheets and mountain glaciers tabulated by Meier (1984) (MG; GMSL = 0.4 mm year^{-1}). The result labeled GIA1 is the $\dot{J}_2$ prediction obtained with a lower mantle viscosity of 3×10^{21} Pa s. (b) As in frame (a), except the observational constraint is based on a new analysis of uncertainties associated with the 18.6-year tide (see text). The prediction labeled GIA1+MM is the total signal associated with the model GIA1 and a twentieth-century melt history composed of Meier's (1984) sources plus 0.6 mm year^{-1} of GMSL rise from polar ice sheets (total GMSL = 1 mm year^{-1}). All GIA calculations adopt an elastic lithospheric thickness of 100 km and an upper mantle viscosity of 5×10^{20} Pa s.

According to these results, the observational constraint is reconciled by GIA predictions based on models with lower mantle viscosity of either 2×10^{21} Pa s or $\sim 2 \times 10^{23}$ Pa s. In analyses by Peltier and colleagues (Peltier, 1983, 1985; Wu and Peltier, 1984), the higher-viscosity root was ruled out by invoking independent GIA constraints associated with relative sea-level records and/or regional gravity trends over Laurentia and Fennoscandia, and it was concluded that the $\dot{J}_2$ datum "can be used to constrain strongly the viscosity of the deep mantle" (Peltier, 1983, p. 434) to values close to 10^{21} Pa s.

Inferences of mantle viscosity based on GIA observations are generally sensitive to uncertainties in the ice load model. In this particular regard, the $\dot{J}_2$ datum is considered to be relatively robust. The degree-two zonal spherical harmonic has the same sign in high-latitude north and south polar regions (Ivins et al., 1993; Mitrovica and Peltier, 1993; **Chapter 3.08**), and thus, predictions of $\dot{J}_2$ are sensitive to the total excess ice volume at last glacial maximum (LGM). That is, they are insensitive to local ice geometries or the partitioning of excess ice volumes into the Southern Hemisphere and Northern Hemisphere. This point was made explicit in early GIA studies (e.g., Peltier, 1983), where predictions such as those shown in **Figure 1(a)** were shown for progressively larger subsets of the Late Pleistocene ice sheets (e.g., Laurentia, Laurentia plus Fennoscandia, and Laurentia plus Fennoscandia plus Antarctica).

To what extent does the simple Earth model parameterization sampled in **Figure 1(a)** capture the sensitivity of the predictions of $\dot{J}_2$ to variations in the radial viscosity profile? Forward analyses indicate that the predictions are insensitive to variations in lithospheric thickness. However, these predictions also demonstrate that the double root structure in the figure largely disappears (at least for the range of deep mantle viscosity considered on the figure) when the upper mantle viscosity is reduced to 10^{20} Pa s or when the boundary defining the viscosity contrast is moved to 1200 km depth (Mitrovica and Peltier, 1993). In these cases, the predictions above a deep mantle viscosity of 2×10^{22} Pa s tend to plateau or diminish more gradually as this viscosity is increased. The origin of this change was explored by Mitrovica and Peltier (1993) who computed radially varying Fréchet kernels that measure the detailed depth-dependent sensitivity of the predictions to perturbations in the radial viscosity profile. They showed that the trend near the low-viscosity 'solutions' in **Figure 1(a)** reflects a sensitivity to structure near the base of the mantle, above the CMB. In contrast, the trend for $\nu_{\mathrm{LM}} \sim 10^{23}$ Pa s results from a sensitivity to viscosity in the top few hundred kilometers of the lower mantle. Thus, if the upper mantle is weakened to 10^{20} Pa s, or the boundary is moved to 1200 km, a prediction of $\dot{J}_2$ will be insensitive to variations in the deep mantle viscosity above values of $\sim 3 \times 10^{22}$ Pa s.

Since the predictions for $\nu_{LM} < 4 \times 10^{21}$ Pa s in Figure 1(a) appear to be robust relative to changes in both the ice history and the shallow structure of the radial Earth model, can one conclude that these predictions reflect a robust constraint on deep mantle viscosity? For a variety of reasons, the answer is no.

First, the early suggestion that the high-viscosity solution on the figure may be rejected on the basis of independent constraints from sea-level histories and regional gravity observations has been weakened by subsequent results. The mathematical expression for the free-air gravity used in these early GIA analysis was in error (Mitrovica and Peltier, 1989), and, more importantly, it is now clear that mantle convective flow beneath Laurentia and Fennoscandia contributes significantly to the static gravity anomaly in these regions (Peltier et al., 1992; Simons and Hager, 1997; Tamisiea et al., 2007). Moreover, because of the presence of lateral variations in mantle viscosity, and the distinct resolving power of the data sets (Mitrovica and Peltier, 1991, 1993), a constraint on viscosity based upon regional sea-level data sets may not be representative of the bulk average mantle viscosity to which the $\dot{J}_2$ datum is sensitive. One may add, in this regard, that there are a suite of GIA-based inferences that prefer a deep mantle viscosity significantly higher than the lower root in **Figure 1(a)** (Mitrovica and Forte, 1997, 2004; Nakada and Lambeck, 1989).

The recognition that a number of geophysical processes other than GIA contribute to the $\dot{J}_2$ datum provides a second and perhaps more fundamental reason to discount inferences of deep mantle viscosity based on analyses of the kind shown in **Figure 1(a)**. After the Yoder et al. (1983) analysis of LAGEOS data, a series of studies demonstrated that ongoing mass flux from polar ice sheets (Greenland and Antarctica) and mountain glaciers would have a potentially large impact on the low-degree Stokes coefficients (Dickey et al., 2002; Gasperini et al., 1986; Ivins et al., 1993; James and Ivins, 1995, 1997; Marcus et al., 2009; Mitrovica and Peltier, 1993; Peltier, 1988; Sabadini et al., 1988; Yoder and Ivins, 1985, Yuen et al., 1987). In **Figure 1(a)**, we show at the top left the size of the $\dot{J}_2$ signal associated with a net mass loss from the Greenland and Antarctic ice sheets equivalent to 1 mm year^{-1} of global mean sea-level (GMSL) rise (as discussed earlier, this degree-2 datum is only sensitive to the net change in polar ice mass), as well as the signal associated with melting from Meier's (1984) tabulation of small glaciers (GMSL$\sim$0.4 mm year). The latter must only be considered as indicative, however, as recent estimates of mass loss from glaciers are larger (e.g., Cogley, 2009; Leclercq et al., 2011) and the relative distribution of these mass losses impacts the contribution to the $\dot{J}_2$ signal (Morrow et al., 2013). Regardless, it is clear from the figure that the signal from ongoing melting may be comparable to the variation with ν_{LM} of the GIA predictions.

In addition to ongoing melting, other analyses have discussed the potential contribution to the $\dot{J}_2$ coefficient from coupling processes at the core–mantle boundary (Dumberry and Bloxham, 2006; Fang et al., 1996), mountain building (Vermeersen et al., 1994), anthropogenic water storage (Chao, 1995), earthquakes (Chao and Gross, 1987), and global water redistribution (Cheng and Tapley 2004; Lavallée et al., 2010; Marcus et al., 2009).

3.07.3 Assessment of the Long-Wavelength Gravity Trends

Since a variety of processes other than GIA contribute non-negligibly to secular trends in the long-wavelength gravity coefficients, the analysis of these harmonics can take several forms. For example, if mantle viscosity is assumed known, then the trends, after correction for GIA, may be used to place bounds on recent melting from ice reservoirs. The latter is a scheme advocated, for example, by Mitrovica and Peltier (1993), who demonstrated that the $\dot{J}_2$ signal due to ongoing polar melting (as well as other low-degree even harmonics; see succeeding text) is relatively insensitive to the geometry of the assumed melting and is linearly proportional to the equivalent GMSL rise associated with the mass flux. As an alternative approach, if one could correct for non-GIA contributions, then the residual could be used to infer deep mantle viscosity. For example, Johnston and Lambeck (1999) argued that a recent nonsteric sea-level rise of 1 mm year^{-1} would imply a lower mantle viscosity close to 10^{22} Pa s. This argument is consistent with the predictions shown in **Figure 1(a)**. Specifically, summing the signals associated with a GIA prediction based on $\nu_{LM} = 10^{22}$ Pa s, melting from Meier's (1984) sources (GMSL = 0.4 mm year^{-1}), and 0.6 mm year^{-1} GMSL rise from ongoing melting of polar ice sheets, yields a total $\dot{J}_2$ prediction of $(-6.0 + 0.9 + 4.0 \times 0.6 = -2.7) \times 10^{-11}$year^{-1}, in accord with the observational constraint.

Either of these approaches, or indeed a simultaneous estimate of viscosity and remnant melting, can also incorporate a suite of long-wavelength harmonics inferred from satellite gravity missions (e.g., Cazenave et al., 1996; Cheng et al., 1989, 1997; Nerem and Klosko, 1996). In this regard, James and Ivins (1997) used secular trends in the zonal harmonics up to degree 4, as well as present-day true polar wander (TPW) rates, to map out a relationship between the inferred deep mantle viscosity and GMSL associated with ongoing melting. They noted, as in the earlier study by Mitrovica and Peltier (1993), both an insensitivity of the low-degree $\dot{J}_l$ predictions to the geometry of ongoing mass flux and a linearity between this signal and the equivalent GMSL value. They inferred a pair of viscosity solutions that tended to merge as the GMSL value rose. This result also has a simple explanation in reference to **Figure 1(a)**. We can interpret the solid line in **Figure 1(a)** as the total signal associated with GIA (versus ν_{LM}) plus a present-day melting from polar ice sheets of GMSL = 0.0 mm year^{-1}. If one increases the melting rate above zero, the total signal will be the same inverted parabola, but shifted downward. The shift will increase as the GMSL rise associated with the present-day melting increases, and thus, the two (low- and high-)viscosity roots that satisfy the observational constraint will eventually merge (for an GMSL rise of $\sim$1.0–1.5 mm year^{-1}) at $\sim 2 \times 10^{22}$ Pa s.

In a related effort, Devoti et al. (2001) (see also Vermeersen et al., 1998) used low-degree zonal harmonics ($\dot{J}_l$, for $l < 7$) that they estimated from LAGEOS, Starlette, and Stella satellite data to infer upper mantle viscosity and lithosphere thickness. They argued that residuals between the satellite-derived observations and best-fit GIA predictions were suggestive of ongoing melting from polar ice sheets. Their analysis was extended by Sabadini et al. (2002) to include the TPW datum. They inferred an increase in viscosity from the upper mantle to the lower

mantle (the latter with a value generally $\sim 10^{22}$ Pa s) and an ongoing melting of the Antarctic ice sheet equivalent to a global mean sea-level rise of 0.7 mm year^{-1}.

The extension to include higher-order harmonics comes at the price of an increasing sensitivity to errors in the Late Pleistocene and present-day ice melt geometries. The solid lines in Figure 2 show predictions of $\dot{J}_l$ due to GIA, for $l = 2$–4, as a function of lower mantle viscosity, computed as in Figure 1(a). The dashed lines are based on a simplified disk load representation of the ICE-3G deglaciation model, as discussed in Mitrovica and Peltier (1993). The even-degree harmonics show larger amplitudes than the odd harmonic $\dot{J}_3$ since the former have the same sign in northern and southern high latitudes (as discussed earlier in the case of degree two), while the latter do not. Therefore, the contributions to $\dot{J}_3$ from the response to loads within these two regions will tend to cancel out (James and Ivins, 1997; Mitrovica and Peltier, 1993). Accordingly, small changes in the geometry of the loads can lead to relatively large changes in the $\dot{J}_3$ predictions, including the change in sign evident in Figure 2(b) as one moves from the ICE-3G to the disk model. One should note also that the sensitivity to changes in the load also increases as one moves from degree two to four, as one would expect since the latter samples at a higher spatial resolution.

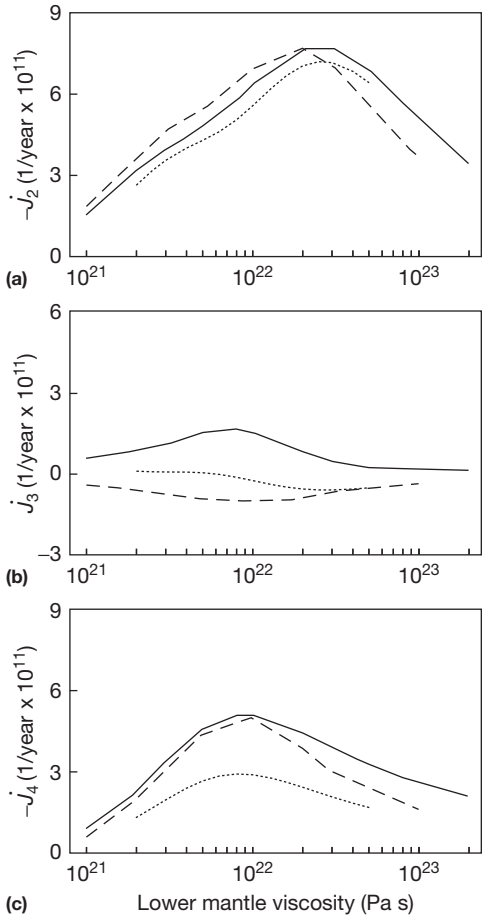


Figure 2 Solid lines – predictions of (a) $\dot{J}_2$, (b) $\dot{J}_3$, and (c) $\dot{J}_4$ due to GIA as a function of the lower mantle viscosity (bottom abscissa scale) of the Earth model. As in Figure 1, the predictions shown by the solid lines use a global Late Pleistocene ice history based on the ICE-3G model (Tushingham and Peltier, 1991). The dashed line on each frame is the analogous prediction based on the disk load ice model discussed in the text. The dotted lines represent predictions using a modified version of the ICE-5G ice history (Peltier, 2004). All calculations adopt an upper mantle viscosity of 5×10^{20} Pa s, and the solid, dashed, and dotted lines adopt lithospheric thicknesses of 100, 100, and 96 km.

Recent deglaciation models have greatly reduced the ice volume loss from Antarctica during LGM (e.g., Ivins and James, 2005; Whitehouse et al., 2012), reducing the magnitude of the region's contribution to the low-degree harmonics (e.g., Simon et al., 2010). However, the decreased contribution from Antarctica will likely result in an increased contribution from other sources, probably in the Northern Hemisphere, in order to satisfy the far-field constraints of total ocean-volume change. To further demonstrate the principles discussed earlier and to illustrate the effect this redistribution of ice may have, the dotted lines in Figure 2 are based on a modified version of ICE-5G (Peltier, 2004), as described in Tamisiea (2011). One of the significant differences between ICE-3G and ICE-5G is the decrease of ice volume in Antarctica, as measured by global average sea-level change, at LGM relative to present day (26 versus 17.3 m) and an increase in Laurentia (58 versus 74 m; Peltier, 2004; Tushingham and Peltier, 1991). Note that the ICE-5G contribution of 17.3 m is still approximately twice the contribution estimated in recent deglaciation models (e.g., Shepard et al., 2012). There is only a slight decrease in the amplitude of $\dot{J}_2$ moving from ICE-3G to ICE-5G. However, the amplitude of $\dot{J}_4$ shows a larger decrease and $\dot{J}_3$ shows a more complex behavior as a function of lower mantle viscosity.

Morrow et al. (2013) had argued that the $\dot{J}_4$ harmonic has significant advantages over the $\dot{J}_2$ harmonic in estimating the recent mass balance of polar ice sheets. For example, the GIA contribution to $\dot{J}_4$ is controlled largely by the history of Antarctic melting since LGM. Moreover, the harmonic is less sensitive than $\dot{J}_2$ to a variety of other geophysical processes (e.g., core–mantle coupling and body tides). Using new inferences of the excess ice volume in the Antarctic at LGM, they demonstrated that the GIA-corrected $\dot{J}_4$ trend over the period 2000–11 inferred from satellite altimetry implies a polar ice sheet and mountain glacier mass balance that is consistent with independent estimates of ice mass loss over the same period (Shepherd et al., 2012).

The consideration of harmonics other than $\dot{J}_2$ can also include trends in the low-degree, nonzonal, or tesseral harmonics of the Earth's gravity field ($\dot{C}_{lm}$ and $\dot{S}_{lm}$ in eqn [1]). However, the earlier lack of observational constraints on these coefficients has meant that the potential GIA signal in these harmonics has received much less attention. Exceptions include studies by Yuen et al. (1987), Ivins et al. (1993), and Tamisiea et al. (2002). The latter two studies demonstrated that these harmonics, which reflect an azimuthal dependence in the gravity trends, are particularly sensitive to uncertainties in the geometry of the Late Pleistocene ice sheets.

3.07.4 Degree-Two Trends in Gravity Revisited

While the effort to consider higher-degree harmonics of the Earth's geopotential has moved forward, recent insights into

both the analysis of the satellite gravity data and the theoretical framework for performing GIA predictions have led to renewed interest in the degree-two trends. In this section, we discuss these issues in turn, beginning with questions that have arisen concerning the error bounds on the observed $\dot{J}_2$.

3.07.4.1 Uncertainties in the $\dot{J}_2$ Constraint

One of the primary sources of error in the estimate of the secular trend in J_2, as determined from satellite gravity data, is the uncertainty in the signal associated with the amplitude and phase of the 18.6-year tide. Cox et al. (2003) had suggested that the latter has been significantly underestimated and that the error bars cited for $\dot{J}_2$ (e.g., the width of the shaded region in **Figure 1(a)**) do not reflect the true uncertainty in the long-term trend. Mitrovica et al. (2006) examined this issue using monthly J_2 values estimated from satellite laser ranging (SLR) data by either Cox and Chao (2002) or Cheng and Tapley (2004). In particular, they corrected these two time series using bounds on the 18.6-year tide signal determined by Benjamin et al. (2006) from the same SLR data sets. They then estimated secular trends from the residual J_2 values over a range of time windows, all beginning in 1979 and ending between 1994 and ~2004. The $\dot{J}_2$ estimates obtained by Mitrovica et al. (2006) did not converge and they were different for the Cox and Chao (2002) and Cheng and Tapley (2004) time series, suggesting that the time series were still too short to accurately correct for decadal fluctuations in J_2. Their results suggested that a more realistic bound on $\dot{J}_2$ was -1.0 to -3.0×10^{-11} year^{-1}, in accord with the conclusion of Cox et al. (2003). This revised uncertainty is shown in **Figure 1(b)**.

The new lower bound for the amplitude of $\dot{J}_2$ clearly permits a higher contribution to the datum from ongoing melting. As an example, consider a model GIA1, defined by a uniform lower mantle of 3×10^{21} Pa s. If this model is accurate, then the observational constraint in **Figure 1(a)** would not permit any significant ongoing melting of ice reservoirs beyond the mass flux (GMSL = 0.4 mm year^{-1}) from Meier's (1984) sources. In contrast, the constraint in **Figure 1(b)** would allow an additional ~0.6 mm year^{-1} GMSL rise from (net) melting of the Antarctic and Greenland ice mass loss (see 'GIA1 + MM' on **Figure 1(b)**). Similar arguments could be made for Earth models with $\nu_{LM} \sim 10^{23}$ Pa s (Mitrovica et al., 2006).

The weakening of the $\dot{J}_2$ constraint provides a potential route to reconciling the so-called 'enigma' of global sea-level rise described by Munk (2002). Munk (2002) recognized that the traditional constraint on $\dot{J}_2$ ($\sim -3 \times 10^{-11}$ year^{-1}; **Figure 1(a)**) was consistent with the change in the rotation rate over the last 2–3 millennia inferred from ancient eclipse observations (Stephenson and Morrison, 1995), and in this case, there could be no anomalous melting of ice reservoirs beginning within the twentieth century. The new lower bound on the $\dot{J}_2$ amplitude introduces an inconsistency with the eclipse inference and allows for the possibility of a significant melting event that onset within the last century (Mitrovica et al., 2006).

The independent constraint on rotation implied by the eclipse records also has implications for previous efforts to reconcile the old $\dot{J}_2$ constraint using GIA models with $\nu_{LM} \sim 10^{22}$ Pa s and ongoing melting of order 1 mm year^{-1} (e.g., James and Ivins, 1997; Johnston and Lambeck, 1999). As discussed, these models predict a net $\dot{J}_2$ signal of $\sim -3 \times 10^{-11}$ year^{-1}, in accordance with **Figure 1(a)**. However, the same models predict a GIA signal of $\sim -6 \times 10^{-11}$ year^{-1} (**Figure 1(a)** for $\nu_{LM} \sim 10^{22}$ Pa s) that would misfit the integrated time shift over the last 2–3 ky inferred from the eclipse records.

As the time span of the data has become longer, a second major uncertainty has become apparent; the J_2 time series is no longer well explained by a simple trend (e.g., Cheng and Tapley, 2004; Cheng et al., 2013; Cox and Chao, 2002). Beginning in the 1990s, the magnitude of the negative trend in the time series significantly decreased (Nerem and Wahr, 2011; Roy and Peltier, 2011). Nerem and Wahr (2011) found that removing a GIA estimate of $\dot{J}_2$ similar to GIA1 in **Figure 1(a)** from the J_2 time series resulted in a nearly flat time series prior to the 1990s and an increasingly positive value after this. They then used GRACE data to estimate the net contribution of present-day mass loss from Greenland and Antarctica to J_2 and found that these estimates agreed well with the post-1990s increase in the residual time series. Uncertainties in GIA contribution and interannual variability (Lavallée et al., 2010) limited their ability to specify the exact timing of the positive trend in the residual time series (Nerem and Wahr, 2011).

3.07.4.2 Geoid Height Change: The Influence of Rotational Feedback

As discussed earlier, changes in the surface mass load associated with the Late Pleistocene glacial cycles perturb the magnitude and orientation of the rotation vector. The former is proportional to the $\dot{J}_2$ datum, while we have denoted the latter as TPW. Since the 1990s, there has been a growing appreciation (Han and Wahr, 1989; Milne and Mitrovica, 1996, 1998; Mitrovica et al., 2001; Peltier, 1998, 2004) that the time-varying centrifugal potential implied by TPW will act as a load that will perturb the gravitational and deformational state of the planet. A cartoon illustrating the geometry of this so-called rotational feedback is shown in **Figure 3**.

The figure shows the situation for a pole moving counterclockwise. In the case of ongoing GIA, the TPW would be in the direction of Hudson Bay (Sabadini and Peltier, 1981; Wu and Peltier, 1984). The centrifugal potential associated with the initial and subsequent pole position is denoted by the dashed and solid ellipses, respectively, and their difference is known as the rotational driving potential. In this two-dimensional cross section along the TPW path, the driving load is negative in the two quadrants toward which the north and south rotation poles are moving, and it is positive in the remaining two quadrants. In each case, the amplitude of the perturbation is a maximum 45° from the pole position, and it is zero along the equator. The rotational driving potential is defined, in three dimensions, by the difference between two ellipsoids of revolution (or oblate spheroids), and the magnitude is also zero along a great circle that passes through the poles and that is 90° from the great circle associated with the TPW path (i.e., a great circle perpendicular to the page through the rotation axis).

Since GIA perturbs the pole position by ~1° or less, the 'quadrential geometry' that defines the rotational driving

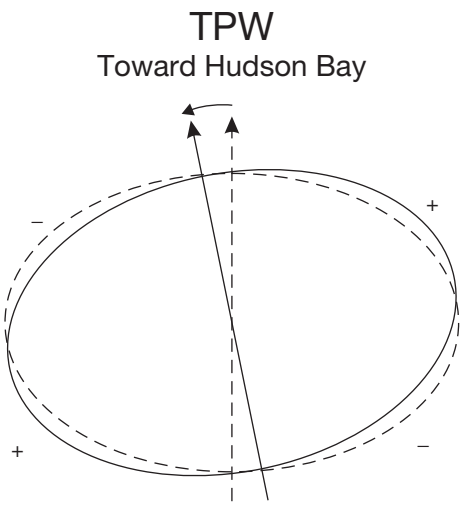


Figure 3 Schematic illustration of the geometry of the GIA-induced rotational feedback signal. The initial rotation pole and associated centrifugal potential are given by the dashed lines. A shift in the rotation pole counterclockwise (or, as labeled on the figure, toward Hudson Bay – as expected in consequence of GIA) leads to a new orientation for both the pole and the centrifugal potential (solid lines). The so-called rotational driving potential is given by the difference between the centrifugal potentials in the two states (solid minus dashed). The direct gravitational effect and deformation driven by this perturbation in the potential will act to bring the geoid (or sea surface) down in the quadrants labeled with '–' and up in quadrants labeled '+.'

potential is strongly dominated by a degree-two, order-one harmonic. Accordingly, the deformational and gravitational response to the driving potential is also dominated by the same harmonic (with the exception of the associated horizontal crustal motions, which have a geometry given by the gradient of the degree-two, order-one surface spherical harmonic; Mitrovica et al., 2001). Indeed, the potentially large imprint of rotational feedback on predictions of GIA-induced rates of change of geoid height (via the perturbation in the degree-two, order-one harmonic) has led to suggestions that this signal should be a target of analysis of data from satellite missions such as GRACE (Peltier, 2004). In this section, we review work that indicates that the feedback signal has been significantly overestimated in some cases, due to both the application of traditional GIA rotation theory and the incorrect use of the resulting predictions.

Figure 4(a) shows a global prediction of the present-day rate of change of sea-surface height due to GIA under the assumption of a nonrotating Earth model. (In the GIA literature, where sea-level changes are computed using an equilibrium, static theory, the equipotential surface corresponding to the ocean's surface is frequently called the 'geoid.' The difference between this definition and meaning associated with 'geoid' in space-geodetic analysis will be discussed later.) The prediction is based on an Earth model with a 90 km thick lithosphere and the VM2 viscosity profile together with a modified version of ICE-5G v1.2 (Peltier, 2004). This modification assumes no prescribed changes in the ice sheets after 4 ky before present (as detailed in Tamisiea, 2011) and six consecutive glaciation and deglaciation cycles. The map is characterized by upwarping over previously glaciated regions at rates that exceed a 1 mm year^{-1} over Laurentia. Outside of these zones, there is a broad downwarping at a rate of $\sim$0.2 mm year^{-1}. The latter trend is associated with the movement of water from the far-field oceans toward the subsiding solid surface at both the periphery of the ancient ice centers and offshore from continents (see Mitrovica and Milne, 2002 for a detailed review of this physics).

In Figure 4(b), we show a prediction analogous to Figure 4(a), with the exception that the signal due to rotational feedback has been included. The latter signal is computed using equations in Milne and Mitrovica (1996) and Mitrovica et al. (2001) (i.e., equations governing the response to the perturbed centrifugal potential associated with a given TPW) and a time-domain version (Mitrovica and Wahr, 2011) of the 'traditional' equations governing GIA-induced TPW derived by Wu and Peltier (1984). (We note that the feedback equations derived by Peltier (1998) include an error in sign; this error, and the associated, predictions were later corrected in Douglas and Peltier (2002) and Peltier (2004).) The large degree-two, order-one quadrential geometry associated with rotational feedback is clearly evident in Figure 4(b), and the orientation of the rotational signature is as one would expect from Figure 3.

Mitrovica et al. (2005) first demonstrated that the traditional equations governing the rotational stability of a viscoelastic planet in response to surface loading, as derived by Wu and Peltier (1984), include an inaccurate assumption that acts to strongly destabilize the rotation pole. The primary issue is the appropriate background form of the Earth upon which the perturbation to the pole position caused by the glacial cycles is computed. The form used by the traditional theory underestimates the true background ellipticity of the Earth and leads to a significant overestimation of the size of the signal due to rotational feedback. Subsequent studies (Cambiotti et al., 2010; Mitrovica and Wahr, 2011; Nakada, 2009; Peltier and Luthcke, 2009) have taken different approaches to the problem, but all agree that a more realistic estimation of the background shape of the Earth leads to a significantly decreased TPW feedback. (It should be noted that Peltier and Luthcke (2009) still advocated the use of the traditional theory.) The basic physics underlying the old rotation theory and a revised formulation derived by Mitrovica et al. (2005) are summarized, schematically, in Figure 5(a) and 5(b).

Any theory governing the rotational stability of the Earth in response to the ice age cycles requires an expression for the response of the model to loading associated with surface masses and the rotational driving potential, as well as an expression for the background oblateness of the model. Both the old and new rotation theories adopt the same Love number formulation for treating the response of a 1-D Earth model to loading, but they diverge in their treatment of the background oblateness. In particular, the old theory assumes that the background oblateness of the Earth can be accurately predicted using the same Love number formulation used to compute the response to the loading. As illustrated in Figure 5(a), this assumption leads to an expression for the oblateness that is a function of the adopted thickness of the elastic lithosphere; the thicker the lithosphere (denoted by the blue outer shell on the top row of the figure), the less oblate the background form (as illustrated by the reduced flattening of frame A2 relative to

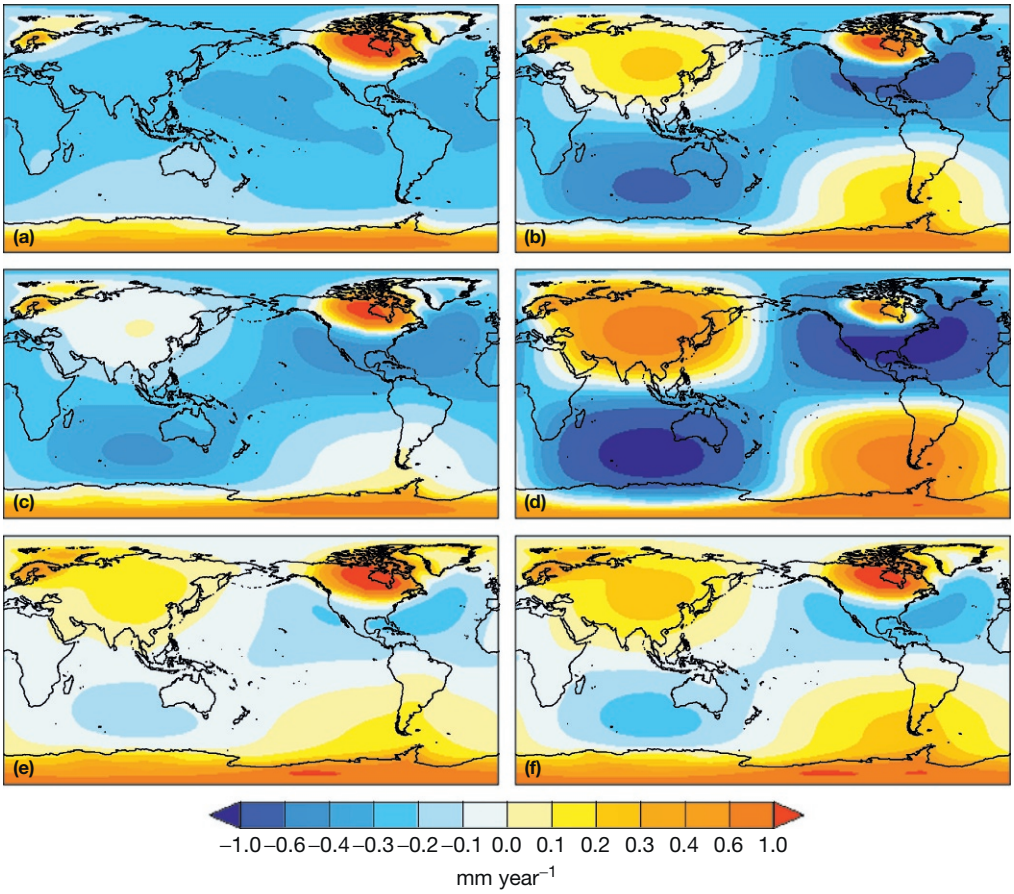


Figure 4 Various GIA predictions computed using an extended version of the ICE-5G ice history and an Earth model with a lithospheric thickness of 90 km and the VM2 viscosity profile (Peltier, 2004). The first three panels show predictions of sea-surface change as would be measured by satellite altimetry; they are distinguished on the basis of the treatment of the rotational feedback due to TPW: (a) no feedback (i.e., a nonrotating Earth model), (b) feedback computed using the traditional GIA rotation theory (e.g., Wu and Peltier, 1984), and (c) feedback computed using the new GIA rotation theory (Mitrovica et al., 2005). (d) As in (b), but with the added effect of an error that increases the degree-two, order-one spherical harmonic coefficients by a factor of $\approx$2.51 (see text). The last two predictions show the geoid change, as would be measured by GRACE, using the feedback computed using (e) the new GIA rotation theory and (f) the traditional theory.

frame A1). The background oblateness of the Earth, that is, the shape upon which ice age forcings are superimposed, should not be a function of the Earth model adopted for the loading calculations. Rather, the observed oblateness is largely a sum of two parts: a hydrostatic form associated with the current rotation rate (B1) and an excess ellipticity due to the dynamic effects of internal convective flow (Nakiboglu, 1982) – as denoted by the red swirls in frame B2. The new rotation theory derived by Mitrovica et al. (2005) incorporates an improved expression for the background oblateness that is tied to the observed form; since the rotation pole is stabilized by the rotational bulge, the old theory, which underestimates the background oblateness, also underestimates the stability of the pole.

As an example, **Figure 5(c)** shows a prediction of the present-day rate of TPW due to GIA as a function of ν_{LM} computed using the old (dotted line) and new (solid line) rotation theory and a simplified loading history (for ICE-5G results, see Mitrovica et al., 2006). The new theory generates a significantly more stable pole, particularly for Earth models with $\nu_{LM} < 10^{22}$ Pa s, and removes the rather enigmatic nonmonotonic sensitivity to ν_{LM} evident in many previous GIA studies of TPW rate. Note that the rate of TPW computed for the model with $\nu_{LM} = 2 \times 10^{21}$ Pa s is reduced by 70% in the new theory relative to the old.

This stabilization is evident in sea-surface rate map in **Figure 4(c)**, where the rotational feedback is computed using the new TPW theory. The viscosity profile of the lower mantle in the VM2 model never reaches a value as large as 4×10^{21} Pa s. Thus, it is in the range of mantle viscosity where the TPW feedback is significantly stabilized in adopting the more accurate rotational theory. Note that the imprint of the feedback mechanism is significantly more muted.

As an example, consider a site located at (105°E, 45°S). The prediction for the nonrotating model (**Figure 4(a)**) yields a rate of -0.17 mm year^{-1}. In contrast, the rate predicted on the basis of the old rotation theory is over three times this value, -0.65 mm year^{-1}. The new rotation theory predicts a subsidence rate of -0.42 mm year^{-1}. The TPW rate associated with the results displayed in **Figure 4(b)** and **4(c)** reduces from 1.0 to 0.39°Ma^{-1}. Clearly, better accounting of the initial form of the Earth, which the glacial cycles perturb, yields a more muted feedback into the sea-surface rate field.

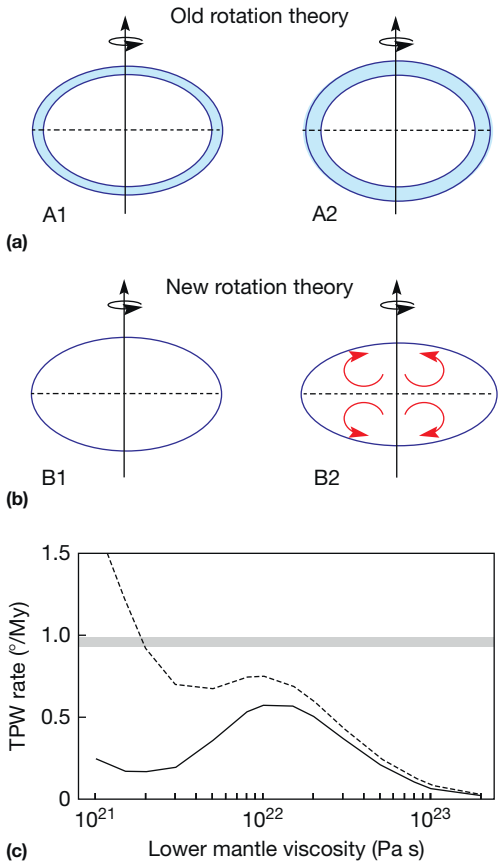


Figure 5 (a)–(b) Schematic illustration of the difference in the underlying choice for the background oblateness of the Earth made within the (a) traditional (e.g., Wu and Peltier, 1984) and (b) new (Mitrovica et al., 2005) GIA rotation theories (see text). (c) Predicted present-day true polar wander rate due to GIA as a function of the lower mantle viscosity of the Earth model. The dotted and solid lines are generated using the traditional and new rotation theories, respectively. All calculations adopt an elastic lithospheric thickness of 100 km and an upper mantle viscosity of 10^{21} Pa s. The shaded region represents an observational constraint on the TPW rate from Gross and Vondrak (1999).

The revision in the TPW rate predictions based on the new rotation theory also has implications for the sea-level enigma mentioned earlier (Munk, 2002). In particular, a model with $\nu_{LM} = 2 \times 10^{21}$ Pa s, which reconciled the old constraint on the $\dot{J}_2$ datum (**Figure 1(a)**), also fits the observational constraint on the TPW rate (**Figure 5(c)**) when the latter prediction is based on the old rotation theory (Wu and Peltier, 1984). However, this fit to the TPW rate disappears when the new theory (Mitrovica et al., 2005) is invoked; therefore, if a viscosity model of this class is correct, then a signal from ongoing melting of ice reservoirs is required to fit the TPW observation.

We noted earlier that many previous efforts to simultaneously fit viscosity and ongoing melt rates using long-wavelength harmonics $\dot{J}_l$ and TPW (e.g., James and Ivins, 1997; Johnston and Lambeck, 1999) had neglected the constraint on integrated rotation rate provided by ancient eclipse data; these studies would also be modified by the new rotation theory that, at the least, would alter the inferred contributions from the Greenland and the Antarctic ice sheets to any ongoing sea-level rise – mass flux from an area further from the rotation pole is a more effective driver of TPW than mass flux from an area closer to it.

Finally, note that whereas the sea-surface prediction shown in **Figure 4(a)–4(c)** is appropriate for use in altimetry data analysis, these fields are not the 'geoid' change estimated in the analysis of GRACE data. **Figure 4(e)** and **4(f)** shows predictions of the GIA contribution to the GRACE measurement for the new and old rotation theory, respectively. These differ from the sea-surface predictions in two ways (Tamisiea, 2011). First, the sea-surface rate predictions have a globally constant rate (a degree-zero term). As the crust under the oceans deforms due to GIA, changing the ocean-basin volume, the sea surface must shift from one equipotential value to another in order for the total ocean volume to be conserved. This globally constant rate term then contributes significantly to the average lowering of the sea surface. Including a globally constant decrease in the geoid, however, would imply that the entire Earth was losing mass, contrary to the assumption used to generate the predictions. The second difference affects the degree-two, order-one coefficients. Because the oceans are in the rotating reference frame, they experience the direct effect of the changes to the centrifugal potential that result from changes in the Earth's rotation. However, the GRACE satellites do not. Deleting this direct centrifugal term, as required in the analysis of GRACE data, leads to a reduction of the amplitude of the degree-two, order-one coefficients by a factor of $(1 + 1/k_f) \approx 2.06$, where k_f is the degree-two fluid Love number of the Earth (Peltier et al., 2012; Tamisiea, 2011).

Using the GIA sea-surface predictions in place of the geoid predictions in a GRACE analysis for the mass gain over the oceans introduces a significant error (as discussed in detail by Chambers et al., 2010). Initial confusion regarding the size of the GIA contribution was further exacerbated by a software bug that leads to the degree-two, order-one coefficients to be amplified by an additional factor of $4\pi/5 \approx 2.51$ in some predictions of the ICE-5G/VM2 model results (e.g., results in Peltier, 2009; Peltier and Lutchke, 2009; error described in Peltier et al., 2012, Text S3; Chambers et al., 2012). The effect of this bug is demonstrated in **Figure 4(d)**, which differs from **Figure 4(b)** only by the factor of 2.51 in the degree-two, order-one coefficients. In analyses of GRACE data, studies often convert geoid changes into change of water thickness that could produce the equivalent geoid change. If this is done for the results in **Figure 4(d)**, **4(e)**, and **4(f)**, the apparent mass loss averaged over the oceans would be 1.81, 0.96, and 1.01 mm year^{-1}, respectively. Thus, the combination of incorrectly using a sea-surface prediction and the software bug (**Figure 4(d)**) leads to a prediction nearly twice as large as the result using the appropriate geoid prediction for the model (**Figure 4(e)**). It should be noted that geoid predictions derived using the old and new theories give very similar results for this contribution, differing by only 0.04 mm year^{-1}. These results for the apparent water thickness change over the oceans due to GIA should be considered indicative only, in that they are highly dependent upon the exact analysis applied to the GRACE data (Chambers et al., 2010; Tamisiea, 2011).

3.07.5 The Influence of Lateral Variations in Mantle Viscosity

The vast majority of published predictions of GIA, including those discussed earlier, have been based on calculations in which the parameters governing the viscoelastic structure of the Earth model vary with depth alone. A set of studies have begun to explore the impact of lateral variations in mantle viscosity on observations of GIA (e.g., Latychev et al., 2005a; Martinec, 2000; Wu and van der Wal, 2003; Zhong et al., 2003). Recently, more studies have focused on regional analysis of GRACE data in areas where GIA is one of the dominant contributors to geoid change (e.g., Steffen et al., 2010, 2012). However, only a few studies have focused on long-wavelength secular trends in the geopotential (e.g., Latychev et al., 2005b, Wu and Wang, 2006).

The Latychev et al. (2005b) results were based on a finite volume formulation of the equations governing the deformational and gravitational response of a Maxwell viscoelastic Earth model. The numerical model was benchmarked against a suite of predictions generated using spherically symmetrical Earth models, including low-degree $\dot{J}_l$ predictions (Latychev et al., 2005a). The input 3-D mantle viscosity field was derived from a model of mantle shear wave heterogeneity (Ritsema et al., 1999) in a three-step process:

(1) The shear wave model was converted to a model for density using a velocity-to-density scaling profile derived by Forte and Woodward (1997).
(2) The density field was mapped into variations in temperature, δT, using a depth-dependent value for the coefficient of thermal expansion (Chopelas and Boehler, 1992).
(3) Temperature was converted to viscosity using the following simple equation:

$$v(\theta, \phi, r) = v_0(r) e^{-\varepsilon \delta T(\theta, \phi, r)} \quad [3]$$

where ε is a free parameter used to vary the strength of the lateral viscosity variations and v_0 is a radially dependent scaling that ensures that the mean (logarithmic) value of the viscosity field at any depth is the same as the spherically symmetrical model to which the 3-D results will be compared.

We will adopt, for the calculations in this section, a background radial model defined by a lithospheric thickness of 100 km, an upper mantle viscosity of 5×10^{20} Pa s, and a lower mantle viscosity of 5×10^{21} Pa s (henceforth model GIA2). As an illustration of the results, we investigate the impact on the GIA2 predictions of introducing heterogeneity in the lower mantle alone. Since we will largely be concerned with the longest-wavelength zonal harmonics, $\dot{J}_l$ for $l = 2$–4, lateral variations in upper mantle viscosity or lithospheric thickness should have less impact on the predictions. Following Latychev et al. (2005b), we consider a suite of results in which ε is varied from 0.0 to 0.06 K^{-1}. For each of these ε values, we computed a histogram of the logarithmic perturbation in viscosity in each lower mantle node relative to the background value of model GIA2 (5×10^{21} Pa s); the plus/minus one standard deviation range for ε values of 0.015, 0.03, and 0.06 K^{-1}, for example, indicated a spread of $\sim$1.0, 1.5, and 3.0 orders of magnitude, respectively (Latychev et al., 2005b). The peak-to-peak range in viscosity was much higher.

As an example of the viscosity field, we show, in **Figure 6**, the logarithmic perturbation in viscosity at a depth of 2000 km generated using $\varepsilon = 0.055$ K^{-1}. The model shows zones of high deep mantle viscosity beneath the Arctic, North America, and the Antarctic. There are also two large regions of low viscosity that sample the megaplume structures beneath the Pacific and southern Africa (Ritsema et al., 1999), which are thought to drive topographic uplift on the overlying lithosphere.

In **Figure 7**, we reproduce (in the solid lines) the $\dot{J}_l$ predictions versus lower mantle viscosity shown in **Figure 2** (based on the ICE-3G load history). The dashed lines map out the perturbations to these predictions generated by considering a suite of 3-D viscosity models defined by increasing ε values up to 0.06 K^{-1} (top abscissa scale). These calculations indicate that the $\dot{J}_2$ prediction is relatively insensitive to the introduction of 3-D viscosity structure; in contrast, the sensitivity to lateral viscosity variations is much higher for the J_4 datum.

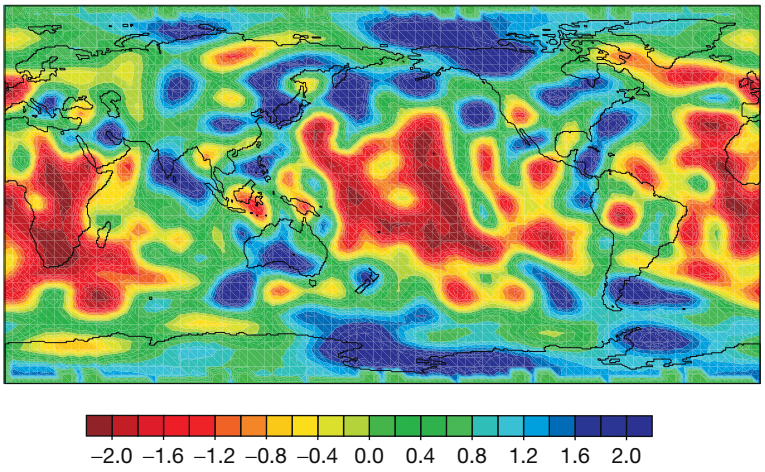


Figure 6 Map of the perturbation in the logarithm of viscosity relative to the mean value associated with model GIA2 (5×10^{21} Pa s) at a depth of 2000 km in the mantle. The field was generated by adopting $\varepsilon = 0.055$ K^{-1} in eqn [3] (see text for details).

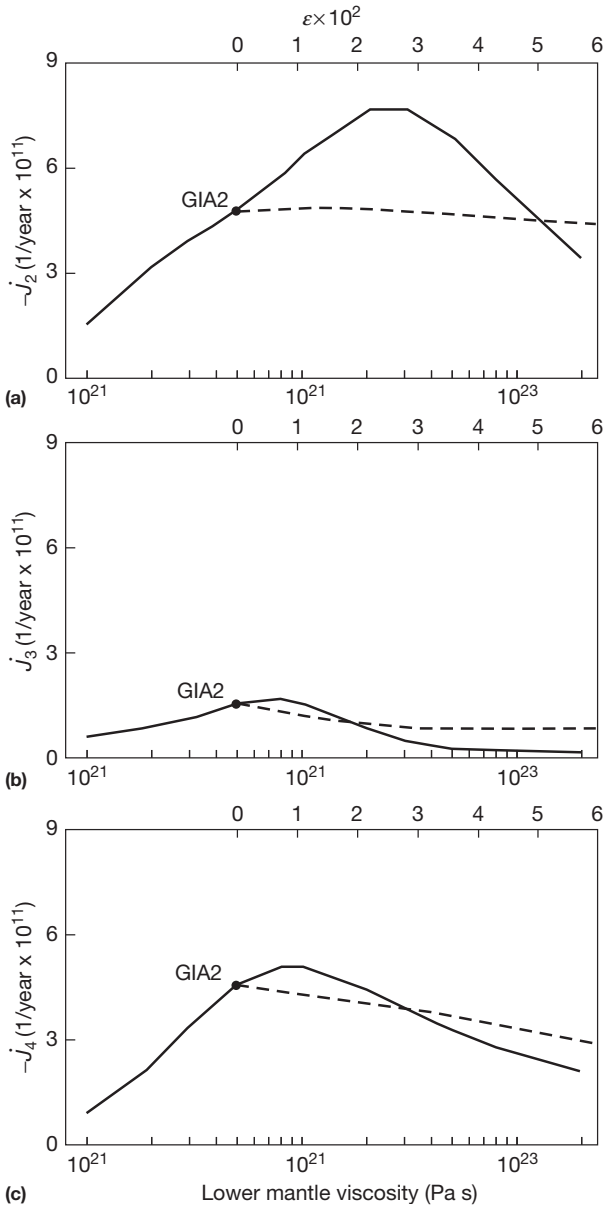


Figure 7 Solid lines – predictions of (a) $\dot{J}_2$, (b) $\dot{J}_3$, and (c) $\dot{J}_4$ due to GIA as a function of the lower mantle viscosity (bottom abscissa scale) of the Earth model (reproduced from **Figure 2**). The dot labeled GIA2 is the prediction obtained with a lower mantle viscosity of 5×10^{21} Pa s. The dashed line on each frame indicates the perturbation in the prediction GIA2 associated with increasing levels of lateral heterogeneity in the lower mantle field, as reflected by the parameter ε (top abscissa scale; see text and eqn [3]). All calculations adopt an elastic lithospheric thickness of 100 km and an upper mantle viscosity of 5×10^{20} Pa s.

The peak perturbation in **Figure 7(c)** is 1.8×10^{-11} year^{-1}, or ~40% reduction relative to the GIA2 prediction of $\dot{J}_4$. This difference is similar to the discrepancy associated with switching the ice history from ICE-3G to ICE-5G in **Figure 2(c)** (for GIA2). However, if the ice sheet history were well known, the differences would be sufficient to significantly bias inferences of deep mantle viscosity based on the assumption of spherical symmetry. Indeed, our 3-D calculation with $\varepsilon = 0.06$ K^{-1} (and an average lower mantle viscosity of 5×10^{21} Pa s) yields the same $\dot{J}_4$ prediction as a 1-D model with $\nu_{LM} = 2.5 \times 10^{21}$ Pa s. The perturbation to $\dot{J}_4$ associated with a net twentieth-century melting from the Greenland and the Antarctic ice sheets is $\sim 3.0 \times 10^{-11}$ year^{-1} (mm year)$^{-1}$ of equivalent GMSL rise (Mitrovica and Peltier, 1993). Thus, the perturbation due to 3-D structure (with $\varepsilon = 0.06$ K^{-1}) is the same as the signal associated with 0.6 mm year^{-1} of ongoing melting from the polar ice complexes. This perturbation is also over an order of magnitude greater than the $\dot{J}_4$ signal associated with Meier's (1984) sources (10^{-12} year^{-1}; Mitrovica and Peltier, 1993) and a factor of 2 larger than the $\dot{J}_4$ signal computed using more recent models of glacier mass balance (Morrow et al., 2013).

Wu and Wang (2006) investigated the effect of lateral heterogeneities on rotational deformation and also concluded that lateral variations in Earth structure had a relatively small impact on the degree-2 coefficients. Lateral variations in earth properties can affect rotational deformation through coupling of spherical harmonic modes and by altering positions of the anomalies relative to the rotation axis. Wu and Wang (2006) found that both of the effects are small, concluding that rotational motion could be approximated by a laterally homogeneous Earth model, which is consistent with the predicted impact on the $\dot{J}_2$ datum.

Finally, in **Figure 8**, we show a map of the difference between predictions of the present-day rate of change of the sea-surface height due to GIA computed using a 3-D Earth model $\varepsilon = 0.055$ K^{-1} and the spherically symmetrical model GIA2. The inclusion of lateral variations in lower mantle viscosity structure acts to reduce the geoid upwarping over Laurentia and the Antarctic by 0.35 and 0.25 mm year^{-1}, respectively, relative to the 1-D model. This reduction appears to be due to the much higher than average deep mantle viscosity beneath these two regions associated with the 3-D model (**Figure 6**). Thus, these high-viscosity zones are acting to retard the relaxation in both high-latitude regions, leading to the reduction in the predicted amplitude of the even zonal harmonics $\dot{J}_2$ and $\dot{J}_4$ in **Figure 7(a)** and **7(c)**.

3.07.6 Summary

The impact of the Late Pleistocene ice age on present-day trends in the long-wavelength gravity field of the Earth has been the focus of continuous study since the first satellite-based estimates of $\dot{J}_2$ were published a quarter century ago. The geophysical applications of these studies have evolved with the growing recognition that a suite of processes may contribute significantly to the observed trends. In particular, early efforts to infer deep mantle viscosity, under the assumption that the observed $\dot{J}_2$ datum was due solely to GIA, have been supplanted by simultaneous estimates of mantle structure and ongoing melting rates based on trends in a set of Stokes coefficients. These studies have commonly been augmented, in some cases, by other GIA observations (e.g., TPW and integrated changes in the length of day inferred from eclipse records).

As a companion to these efforts, theoretical treatments of the impact of GIA on the long-wavelength gravity field have also advanced beyond the earliest calculations based on spherically symmetrical, nonrotating Earth models. In recent years, the feedback of glaciation-induced polar wander has been included in the predictions and these have had to be updated in

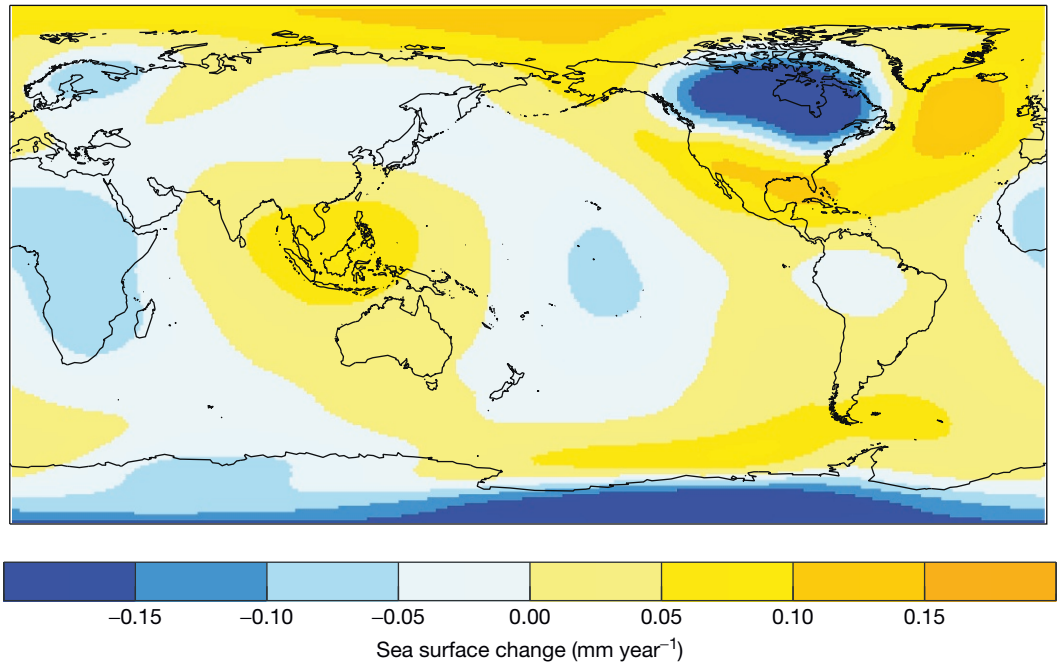


Figure 8 Perturbation in the prediction of the present-day change of the sea-surface height due to GIA associated with the introduction of lateral variations in mantle viscosity within the lower mantle defined by $\varepsilon = 0.055\ \mathrm{K}^{-1}$ (as in **Figure 6**). That is, the prediction for the 3-D model minus the prediction based on the spherically symmetrical model GIA2 (in each case, rotational feedback is not included).

consequence of a significant improvement in the theory governing load-induced perturbations in the rotation vector. Moreover, an extension to Earth models with complex, 3-D viscoelastic structure has also become possible. These calculations indicate that lateral variations in deep mantle viscosity can significantly perturb predictions of the $\dot{J}_l$ coefficients for $l > 2$.

The lessons learned from these studies, and the theoretical advances they motivated, remain important as attention moves increasingly toward the constraints on the Earth's time-variable gravity field provided by the GRACE satellite mission. Indeed, regional studies of gravity trends over Greenland and Antarctica, with the goal of estimating recent mass flux, have been compelled to assess the possible range of GIA signal. In particular, recent models of the ice history of Antarctica have been characterized by significant reductions in the excess ice at the LGM, and this has significantly altered GRACE-based estimates of present-day mass loss from the continent (e.g., Shepherd et al., 2012). In addition, as discussed in **Section 3.07.4.2**, a variety of shortcomings in previous analyses of low-degree components in the rate of change of the geopotential have led to large discrepancies between estimates of the GIA contribution to GRACE observations of the ocean-mass change (e.g., Chambers et al., 2010; Tamisiea, 2011).

References

Alexander JC (1983) Higher harmonic effects of the Earth's gravitational field from postglacial rebound as observed from LAGEOS. *Geophysical Research Letters* 10: 1085–1087.

Argus DF (1996) Postglacial rebound from VLBI geodesy: On establishing vertical reference. *Geophysical Research Letters* 23: 973–977.

Benjamin D, Wahr J, Ray R, Egbert G, and Desai SD (2006) Constraints on mantle anelasticity from geodetic observations, and implications for the J_2 anomaly. *Geophysical Journal International* 165: 316.

Calais E, Han JY, DeMets C, and Nocquet JM (2006) Deformation of the North American plate interior from a decade of continuous GPS measurements. *Journal of Geophysical Research* 111: B06402. http://dx.doi.org/10.1029/2005JB004253.

Cambiotti G, Ricard Y, and Sabadini R (2010) Ice age True Polar Wander in a compressible and non-hydrostatic Earth. *Geophysical Journal International* 183: 1248–1264. http://dx.doi.org/10.1111/j.1365-246X.2010.04791.x.

Cathles LM (1975) *The Viscosity of the Earth's Mantle*. Princeton, NJ: Princeton Univ. Press.

Cazenave AA, Gegout P, Ferhat G, and Biancale R (1996) Temporal variations of the gravity field from LAGEOS 1 and LAGEOS 2 observations. In: Rapp R, Cazenave AA, and Nerem RS (eds.) *Global Gravity Field and Its Temporal Variations*, pp. 141–151. New York: Springer.

Chambers DP, Wahr J, Tamisiea ME, and Nerem RS (2010) Ocean mass from GRACE and glacial isostatic adjustment. *Journal of Geophysical Research* 115: B11415. http://dx.doi.org/10.1029/2010JB007530.

Chambers DP, Wahr J, Tamisiea ME, and Nerem RS (2012) Reply to comment by W. R. Peltier et al. on "Ocean mass from GRACE and glacial isostatic adjustment". *Journal of Geophysical Research* 117, B11404.

Chao BF (1995) Anthropogenic impact on global geodynamics due to reservoir water impoundment. *Geophysical Research Letters* 22: 3529–3532.

Chao BF and Gross RS (1987) Changes in the Earth's rotation and low-degree gravitational field induced by earthquakes. *Geophysical Journal of the Royal Astronomical Society* 91: 569–596.

Cheng MK, Eanes RJ, Shum CK, Schutz BE, and Tapley BD (1989) Temporal variations in low degree zonal harmonics from Starlette orbit analysis. *Journal of Geophysical Research* 16: 393–396.

Cheng MK, Shum CK, and Tapley BD (1997) Determination of long-term changes in the Earth's gravity field from satellite laser ranging observations. *Journal of Geophysical Research* 102: 22377–22390.

Cheng M and Tapley BD (2004) Variations in the Earth's oblateness during the past 28 years. *Journal of Geophysical Research* 109, B09402.

Cheng M, Tapley BD, and Ries JC (2013) Deceleration in the Earth's oblateness. *Journal of Geophysical Research* 118: 740–747. http://dx.doi.org/10.1002/jgrb.50058.

Chopelas A and Boehler R (1992) Thermal expansivity in the lower mantle. *Geophysical Research Letters* 19: 1983–1986.

Cogley JG (2009) Geodetic and direct mass-balance measurements: Comparison and joint analysis. *Annals of Glaciology* 50: 96–100.

Cox CM, Boy A, and Chao BF (2003) Time-variable gravity: using satellite-laser-ranging as a tool for observing long-term changes in the Earth system. In: Noomen R, Klosko S, Noll C, and Pearlman M (eds.) *Proceedings of the 13th International Workshop on Laser Ranging, NASA/CP-2003–212248* 9–19.

Cox CM and Chao BF (2002) Detection of a large-scale mass redistribution in the terrestrial system since 1998. *Science* 297: 831–832.

Devoti R, Luceri V, Sciarretta C, et al. (2001) The SLR secular gravity variations and their impact on the inference of mantle rheology and lithospheric thickness. *Geophysical Research Letters* 28: 855–858.

Dickey JO, Marcus SL, de Viron O, and Fukumori I (2002) Recent earth oblateness variations: Unraveling climate and postglacial rebound effects. *Science* 298: 1975–1977. http://dx.doi.org/10.1126/science.1077777.

Dietrich R, Ivins E, Casassa G, Lange H, Wendt J, and Fritsche M (2010) Rapid crustal uplift in Patagonia due to enhanced ice loss. *Earth and Planetary Science Letters* 289: 22–29. http://dx.doi.org/10.1016/j.epsl.2009.10.021.

Douglas BC and Pelter WR (2002) The puzzle of global sea-level rise. *Physics Today* 55: 3540.

Dumberry M and Bloxham J (2006) Azimuthal flows in the Earth's core and changes in length of day at millennial timescales. *Geophysical Journal International* 16: 32–46.

Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model (PREM). *Physics of Earth and Planetary Interiors* 25: 297–356.

Fang M, Hager BH, and Herring TA (1996) Surface deformation caused by pressure changes in the fluid core. *Geophysical Research Letters* 23: 1493–1496.

Forte AM and Woodward RL (1997) Seismic-geodynamic constraints on three-dimensional structure, vertical flow, and heat transfer in the mantle. *Journal of Geophysical Research* 102: 17981–17994.

Gasperini P, Sabadini R, and Yuen DA (1986) Excitation of the Earth's rotational axis by recent glacial discharges. *Geophysical Research Letters* 13: 533–536.

Gross RS and Vondrak J (1999) Astrometric and space-geodetic observations of polar wander. *Geophysical Research Letters* 26: 2085–2088.

Han D and Wahr J (1989) Post-glacial rebound analysis for a rotating Earth. In: Cohen S and Vanicek P (eds.) *AGU Geophysical Monograph Series 49: Slow Deformations and Transmission of Stress in the Earth*, pp. 1–6. Washington, DC: AGU.

Ivins ER and James TS (2005) Antarctic glacial isostatic adjustment: A new assessment. *Antarctic Science* 17: 541–553. http://dx.doi.org/10.1017/S0954102005002968.

Ivins ER, Sammis CG, and Yoder CF (1993) Deep mantle viscous structure with prior estimate and satellite constraint. *Journal of Geophysical Research* 98: 4579–4609.

James TS and Ivins ER (1995) Present-day Antarctic ice mass changes and crustal motion. *Geophysical Research Letters* 22: 973–976.

James TS and Ivins ER (1997) Global geodetic signatures of the Antarctic ice sheet. *Journal of Geophysical Research* 102: 605–633.

James TS and Lambert A (1993) A comparison of VLBI data with the ICE-3G glacial rebound model. *Geophysical Research Letters* 20: 871–874.

Johansson JM, Davis JL, Scherneck H-G, et al. (2002) Continuous measurements of post-glacial adjustment in Fennoscandia, 1: Geodetic results. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000400.

Johnston P and Lambeck K (1999) Postglacial rebound and sea level contributions to changes in the geoid and the Earth's rotation axis. *Geophysical Journal International* 136: 537–5889.

Latychev K, Mitrovica JX, Tamisiea ME, Tromp J, Christara C, and Moucha R (2005b) GIA-induced secular variations in the Earth's long wavelength gravity field: Influence of 3-D viscosity variations. *Earth and Planetary Science Letters* 240: 322–327.

Latychev K, Mitrovica JX, Tromp J, Tamisiea ME, Komatitsch D, and Christara C (2005a) Glacial isostatic adjustment on 3-D Earth models: A New finite-element formulation. *Geophysical Journal International* 161: 421–444.

Lavallée DA, Moore P, Clarke PJ, Petrie EJ, van Dam T, and King MA (2010) J2: An evaluation of new estimates from GPS, GRACE, and load models compared to SLR. *Geophysical Research Letters* 37: L22403. http://dx.doi.org/10.1029/2010GL045229.

Leclercq PW, Oerlemans J, and Cogley JG (2011) Estimating the glacier contribution to sea-level rise for the period 1800–2005. *Surveys in Geophysics* 32: 519–535. http://dx.doi.org/10.1007/s10712-011-9121-7.

Marcus SL, Dickey JO, Willis JK, and Seitz F (2009) Earth oblateness changes reveal land ice contribution to interannual sea level variability. *Geophysical Research Letters* 36: L23608. http://dx.doi.org/10.1029/2009GL041130.

Martinec Z (2000) Spectral-finite element approach to three-dimensional viscoelastic relaxation in a spherical earth. *Geophysical Journal International* 142: 117–141.

McConnell RK (1968) Viscosity of the mantle from relaxation time spectra of isostatic adjustment. *Journal of Geophysical Research* 73: 7089–7105.

Meier MF (1984) Contributions of small glaciers to global sea level. *Science* 226: 1418–1421.

Milne GA, Davis JL, Mitrovica JX, et al. (2001) Space-geodetic constraints on glacial isostatic adjustment in Fennoscandia. *Science* 291: 2381–2385.

Milne GA and Mitrovica JX (1996) Postglacial sea-level change on a rotating Earth: First results from a gravitationally self-consistent sea-level equation. *Geophysical Journal International* 126: F13–F20.

Milne GA and Mitrovica JX (1998) Postglacial sea-level change on a rotating Earth. *Geophysical Journal International* 133: 1–10.

Mitrovica JX, Davis JL, and Shapiro II (1993) Constraining proposed combinations of ice history and earth rheology using VLBI determined baseline length rates in North America. *Geophysical Research Letters* 20: 2387–2390.

Mitrovica JX, Davis JL, and Shapiro II (1994) A spectral formalism for computing three dimensional deformations due to surface loads—II. Present-day glacial isostatic adjustment. *Journal of Geophysical Research* 99: 7075–7101.

Mitrovica JX and Forte AM (1997) Radial profile of mantle viscosity: Results from the joint inversion of convection and postglacial rebound observables. *Journal of Geophysical Research* 102: 2751–2769.

Mitrovica JX and Forte AM (2004) A New inference of mantle viscosity based upon a joint inversion of convection and glacial isostatic adjustment data. *Earth and Planetary Science Letters* 225: 177–189.

Mitrovica JX and Milne GA (2002) On the origin of postglacial ocean syphoning. *Quaternary Science Reviews* 21: 2179–2190.

Mitrovica JX, Milne GA, and Davis JL (2001) Glacial isostatic adjustment of a rotating Earth. *Geophysical Journal International* 147: 562–578.

Mitrovica JX and Peltier WR (1989) Pleistocene deglaciation and the global gravity field. *Journal of Geophysical Research* 96: 13651–13657.

Mitrovica JX and Peltier WR (1991) A complete formalism for the inversion of post-glacial rebound data: Resolving power analysis. *Geophysical Journal International* 104: 267–288.

Mitrovica JX and Peltier WR (1993) Present-day secular variations in the zonal harmonics of the Earth's geopotential. *Journal of Geophysical Research* 98: 4509–4526.

Mitrovica JX and Wahr J (2011) Ice age earth rotation. *Annual Review of Earth and Planetary Sciences* 39: 577–616.

Mitrovica JX, Wahr J, Matsuyama I, and Paulson A (2005) The rotational stability of an ice age Earth. *Geophysical Journal International* 161: 491–506.

Mitrovica JX, Wahr J, Matsuyama I, and Paulson A (2006) Reanalysis of ancient eclipse, astronomic and geodetic data: A possible route to resolving the enigma of global sea-level rise. *Earth and Planetary Science Letters* 243: 390–399.

Morrow E, Mitrovica JX, Sterenborg MG, and Harig C (2013) A test of recent inferences of NET polar ice mass balance based on long-wavelength gravity. *Journal of Climate e-View*. http://dx.doi.org/10.1175/JCLI-D-13-00078.1.

Munk W (2002) Twentieth century sea level: An enigma. *Proceedings of the National Academy of Sciences* 99: 6550–6555.

Nakada M (2009) Polar wander of the Earth associated with the Quaternary glacial cycle on a convecting mantle. *Geophysical Journal International* 179: 569–578. http://dx.doi.org/10.1111/j.1365-246X.2009.04289.x.

Nakada M and Lambeck K (1989) Late Pleistocene and Holocene sea-level change in the Australian region and mantle rheology. *Geophysical Journal International* 96: 497–517.

Nakiboglu SM (1982) Hydrostatic theory of the Earth and its mechanical implications. *Physics of the Earth and Planetary Interiors* 28: 302–311.

Nakiboglu SM and Lambeck K (1980) Deglaciation effects upon the rotation of the Earth. *Geophysical Journal of the Royal Astronomical Society* 62: 49–58.

Nerem RS and Klosko SM (1996) Secular variations of the zonal harmonics and polar motion as geophysical constraints. In: Rapp R, Cazenave AA, and Nerem RS (eds.) *Global Gravity Field and Its Temporal Variations*, pp. 152–163. New York: Springer.

Nerem RS and Wahr J (2011) Recent changes in the Earth's oblateness driven by Greenland and Antarctic ice mass loss. *Geophysical Research Letters* 38: L13501. http://dx.doi.org/10.1029/2011GL047879.

Peltier WR (1983) Constraint on deep mantle viscosity from LAGEOS acceleration data. *Nature* 304: 434–436.

Peltier WR (1985) The LAGEOS constraint on deep mantle viscosity: Results from a new normal mode method for the inversion of viscoelastic relaxation spectra. *Journal of Geophysical Research* 90: 9411–9421.

Peltier WR (1988) Global sea level and Earth rotation. *Science* 240: 895–901.

Peltier WR (1995) VLBI baselines for the ICE-4G model of postglacial rebound. *Geophysical Research Letters* 22: 465–469.

Peltier WR (1998) Postglacial variations in the level of the sea: Implications for climate dynamics and solid-Earth geophysics. *Reviews of Geophysics* 36: 603–689.

Peltier WR (2004) Global glacial isostasy and the surface of the ice-age Earth: The ICE-5G (VM2) model and GRACE. *Annual Review of Earth and Planetary Sciences* 32: 111–149.

Peltier WR (2009) Closure of the budget of global sea level rise over the GRACE era: The importance and magnitudes of the required corrections for global glacial isostatic adjustment. *Quaternary Science Reviews* 28: 1658–1674. http://dx.doi.org/10.1016/j.quascirev.2009.04.004.

Peltier WR and Andrews JT (1976) Glacial-isostatic adjustment I – The forward problem. *Geophysical Journal of the Royal Astronomical Society* 46: 605–646.

Peltier WR, Drummond R, and Roy K (2012) Comment on "Ocean mass from GRACE and glacial isostatic adjustment" by D. P. Chambers et al. *Journal of Geophysical Research* 117, B11403.

Peltier WR, Forte AM, Mitrovica JX, and Dziewonski AM (1992) Earth's gravitational field: Seismic tomography resolves the enigma of the Laurentian anomaly. *Geophysical Research Letters* 19: 1555–1558.

Peltier WR and Luthcke SB (2009) On the origins of Earth rotation anomalies: New insights on the basis of both "paleogeodetic" data and Gravity Recovery and Climate Experiment (GRACE) data. *Geophysical Research Letters* 114: B11405. http://dx.doi.org/10.1029/2009JB006352.

Ritsema J, van Heijst HJ, and Woodhouse JH (1999) Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science* 286: 1925–1928.

Roy K and Peltier WR (2011) GRACE era secular trends in Earth rotation parameters: A global scale impact of the global warming process? *Geophysical Research Letters* 38: L10306. http://dx.doi.org/10.1029/2011GL047282.

Rubincam D (1984) Postglacial rebound observed by LAGEOS and the effective viscosity of the lower mantle. *Journal of Geophysical Research* 89: 1077–1087.

Sabadini R, DiDonato G, Vermeersen LLA, Devoti R, Luceri V, and Bianco G (2002) Ice mass loss in Antarctica and stiff lower mantle viscosity inferred from the long wavelength time dependent gravity field. *Geophysical Research Letters* 29. http://dx.doi.org/10.1029/2001GL014016.

Sabadini R and Peltier WR (1981) Pleistocene deglaciation and the Earth's rotation: Implications for mantle viscosity. *Geophysical Journal of the Royal Astronomical Society* 66: 553–578.

Sabadini R, Yuen DA, and Gasperini P (1988) Mantle rheology and satellite signatures from present-day glacial forcings. *Journal of Geophysical Research* 12: 437–447.

Sella GF, Stein S, Dixon TH, et al. (2007) Observation of glacial isostatic adjustment in "stable" North America with GPS. *Geophysical Research Letters* 34: L02306. http://dx.doi.org/10.1029/2006GL027081.

Shepherd A, Ivins ER, Geruo A, et al. (2012) Reconciled estimate of Ice-sheet mass balance. *Science* 338: 1183–1189. http://dx.doi.org/10.1126/science.1228102.

Simon K, James T, and Ivins E (2010) Ocean loading effects on the prediction of Antarctic glacial isostatic uplift and gravity rates. *Journal of Geodesy* 84: 305–317. http://dx.doi.org/10.1007/s00190-010-0368-4.

Simons M and Hager BH (1997) Localization of the gravity field and the signature of glacial rebound. *Nature* 390: 500–504.

Steffen H, Wu P, and Wang H (2010) Determination of the Earth's structure in Fennoscandia from GRACE and implications for the optimal post-processing of GRACE data. *Geophysical Journal International* 182: 1295–1310. http://dx.doi.org/10.1111/j.1365-246X.2010.04718.x.

Steffen H, Wu P, and Wang H (2012) Optimal locations for absolute gravity measurements and sensitivity of GRACE observations for constraining glacial isostatic adjustment on the northern hemisphere. *Geophysical Journal International* 190: 1483–1494. http://dx.doi.org/10.1111/j.1365-246X.2012.05563.x.

Stephenson FR and Morrison LV (1995) Long-term fluctuations in the Earth's rotation: 700 BC to AD 1990. *Proceedings of the Royal Society of London A* 351: 165–202.

Tamisiea ME (2011) Ongoing glacial isostatic contributions to observations of sea level change. *Geophysical Journal International* 186: 1036–1044. http://dx.doi.org/10.1111/j.1365-246X.2011.05116.x.

Tamisiea ME, Mitrovica JX, and Davis JL (2007) GRACE gravity data constrain ancient Ice geometries and continental dynamics over Laurentia. *Science* 316: 881–883. http://dx.doi.org/10.1126/science.1137157.

Tamisiea ME, Mitrovica JX, Tromp J, and Milne GA (2002) Present-day secular variations in the low-degree harmonics of the geopotential: Sensitivity analysis on spherically symmetric Earth models. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000696.

Tapley B, Bettadpur S, Ries J, Thompson P, and Watkins M (2004) GRACE measurements of mass variability in the Earth system. *Science* 205: 503–505.

Thomas ID, King MA, Bentley MJ, et al. (2011) Widespread low rates of Antarctic glacial isostatic adjustment revealed by GPS observations. *Geophysical Research Letters* 38: L22302. http://dx.doi.org/10.1029/2011GL049277.

Tushingham AM and Peltier WR (1991) ICE-3G: A new global model of late Pleistocene deglaciation based upon geophysical predictions of post-glacial relative sea level change. *Journal of Geophysical Research* 96: 4497–4523.

Vermeersen LLA, Sabadini R, Devoti R, et al. (1998) Mantle viscosity from joint inversions of Pleistocene deglaciation-induced changes in geopotential with a new SLR analysis and polar wander. *Geophysical Research Letters* 25: 4261–4264.

Vermeersen LLA, Sabadini R, Spada G, and Vlaar NJ (1994) Mountain building and Earth rotation. *Geophysical Journal International* 117: 610–624.

Walcott RI (1972) Late Quaternary vertical movements in Eastern North America: Quantitative evidence of glacio-isostatic rebound. *Reviews of Geophysics and Space Physics* 10: 849–884.

Whitehouse PL, Bentley MJ, and Brocq AML (2012) A deglacial model for Antarctica: Geological constraints and glaciological modelling as a basis for a new model of Antarctic glacial isostatic adjustment. *Quaternary Science Reviews* 32: 1–24. http://dx.doi.org/10.1016/j.quascirev.2011.11.016.

Wu P and Peltier WR (1983) Glacial isostatic adjustment and the free air gravity anomaly as a constraint on deep mantle viscosity. *Geophysical Journal of the Royal Astronomical Society* 74: 377–449.

Wu P and Peltier WR (1984) Pleistocene deglaciation and the Earth's rotation: A new analysis. *Geophysical Journal of the Royal Astronomical Society* 76: 753–791.

Wu P and van der Wal W (2003) Postglacial sealevels on a spherical, self-gravitating viscoelastic Earth: Effects of lateral viscosity variations in the upper mantle on the inference of viscosity contrasts in the lower mantle. *Earth and Planetary Science Letters* 211: 57–68.

Wu P and Wang H (2006) Effects of mode coupling and location of rotational axis on glacial induced rotational deformation in a laterally heterogeneous viscoelastic earth. *Geophysical Journal International* 167: 853–859. http://dx.doi.org/10.1111/j.1365-246X.2006.03103.x.

Yoder CF and Ivins ER (1985) Changes in the Earth's gravity field from Pleistocene deglaciation and present-day glacial melting. *Eos, Transactions American Geophysical Union* 66(18): 245.

Yoder C, Williams J, Dickey J, Schutz B, Eanes R, and Tapley B (1983) Secular variations of the earth's gravitational harmonic J_2 coefficient from Lageos and nontidal acceleration of earth rotation. *Nature* 303: 757–762.

Yuen DA, Gasperini P, Sabadini R, and Boschi E (1987) Azimuthal dependence in the gravity field induced by recent and past cryospheric forcings. *Geophysical Research Letters* 8: 812–815.

Yuen DA and Sabadini R (1985) Viscosity stratification of the lower mantle as inferred from the $\dot{J}_2$ observation. *Annales de Geophysciae* 3: 647–654.

Zhong S, Paulson A, and Wahr J (2003) Three-dimensional finite element modelling of Earth's viscoelastic deformation: Effects of lateral variations in lithospheric thickness. *Geophysical Journal International* 155: 679–695.

This page intentionally left blank

3.08 Time-Variable Gravity from Satellites

JM Wahr, University of Colorado, Boulder, CO, USA

© 2015 Elsevier B.V. All rights reserved.

3.08.1	**Introduction**	193
3.08.1.1	Nonuniqueness	193
3.08.1.2	Time-Variable Gravity	194
3.08.1.3	Changes in the Earth's Oblateness	195
3.08.2	**GRACE**	197
3.08.2.1	Gravity Solutions	197
3.08.2.2	Using the Harmonic Solutions to Solve for Mass	197
3.08.2.3	Love Numbers	198
3.08.2.4	Spatial Averaging	199
3.08.2.4.1	Smoothing	199
3.08.2.4.2	Regional averaging	202
3.08.2.5	Estimating Errors and Accounting for Leakage	202
3.08.2.5.1	Other methods	204
3.08.3	**Applications**	204
3.08.3.1	Hydrology	204
3.08.3.1.1	Comparing with land surface models	205
3.08.3.1.2	Floods and droughts	205
3.08.3.1.3	Anthropogenic effects and sea level contributions	205
3.08.3.1.4	Precipitation (P) minus evapotranspiration (ET)	206
3.08.3.2	Cryosphere	207
3.08.3.3	Solid Earth	208
3.08.3.4	Oceanography	209
3.08.4	**The Future**	210
References		210

3.08.1 Introduction

The Earth's gravity field is a product of its mass distribution, mass both deep within the Earth and at and above its surface. That mass distribution is constantly changing. Tides in the ocean and solid Earth cause large mass variations at 12 and 24 h periods. Atmospheric disturbances associated with synoptic storms and with seasonal and interannual climatic variations lead to variations in the distribution of mass in the atmosphere, the ocean, and the water stored on land. Mantle convection induces mass variations throughout the mantle that have large amplitudes compared to those associated with climatic variability but that generally occur slowly across human timescales.

Because of these and other processes, the Earth's gravity field varies with time. Observations of that variability using either satellites or ground-based instrumentation can be used to study a wide variety of geophysical processes that involve changes in mass (Dickey et al., 1997). Solid Earth processes are not usually the main focus of time-variable gravity measurements. Instead, most of the time-variable signal comes from the Earth's fluid envelope: the oceans, the atmosphere, the polar ice sheets and continental glaciers, and the storage of water and snow on land. Fluids (water and gasses) are much more mobile than rock.

Solid Earth deformation does have a significant indirect effect on ground-based gravity measurements. A gravimeter on the Earth's surface is sensitive to vertical motion of that surface. When the surface goes up, the gravimeter moves further from the center of the Earth and so it sees a smaller gravitational acceleration. For most solid Earth processes, the signal from the vertical displacement of the meter is several times larger than the gravity change caused by the displaced mass. Thus, a surface gravimeter can arguably be viewed as a vertical positioning instrument. A satellite, on the other hand, is not fixed to the surface, and so the gravity signals it detects are due entirely to the underlying mass distribution. Satellite gravity thus provides direct constraints on that mass.

3.08.1.1 Nonuniqueness

One serious limitation when interpreting gravity observations is that the inversion of gravity for density is nonunique. There are always an infinite number of possible internal density distributions that can produce the same external gravity field. Even perfect knowledge of the external gravity field would not provide a unique density solution.

As a simple illustration of this nonuniqueness, consider the gravity field outside a sphere. The external gravitational acceleration is $g = MG/r^2$, where M is the total mass of the sphere, G is Newton's gravitational constant, and r is the distance to the center of the sphere. This same expression holds whether the mass is uniformly distributed throughout the sphere or is localized entirely at the outer surface or has any other radially dependent distribution. By observing the external gravity field in this case, all that could be learned is the total mass of the

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00065-8

sphere and the fact that the internal density is spherically symmetrical. The details of how the density is distributed with radius would remain unknown. This nonuniqueness would disappear if the gravity field everywhere inside the sphere were also known. But knowledge of the external field alone is not enough.

This nonuniqueness is a major limitation when interpreting the Earth's static gravity field. For example, Figure 1 shows a map of the Earth's static geoid anomaly, as determined by Lemoine et al. (1998) from decades of satellite and surface observations. The geoid is the surface of constant potential that coincides with mean sea level over the ocean. The geoid anomaly is the elevation of the geoid above its mean ellipsoidal average. This is a common method of representing the Earth's gravity field, one that emphasizes the long wavelength characteristics of the field. There is a trade-off between amplitude and depth when using this map to constrain the Earth's time-averaged mass distribution. For example, from this map alone, it is not possible to know whether the large red feature over Indonesia is caused by a large positive mass anomaly in the crust or a much larger positive mass anomaly deeper in the mantle.

But Figure 1 clearly does contain information about the Earth's internal density. It represents a weighted vertical average of the underlying mass anomalies. Static gravity observations are particularly useful when combined with independent information or assumptions about the depth of the density anomaly or its spatial pattern.

3.08.1.2 Time-Variable Gravity

Nonuniqueness is much less of an issue for time-varying gravity. Time-varying signals, if they vary rapidly enough, can usually be assumed to come from mass variability at the Earth's surface rather than from deep within the Earth. For example, Figure 2(a) shows the amplitude of the annual cycle in the geoid as observed from the Gravity Recovery and Climate Experiment (GRACE; see succeeding text). It is almost certain that this signal is coming from some combination of the atmosphere, the oceans, and the water/snow/ice stored on or just below the land surface. Few solid Earth processes are likely to vary this rapidly, let alone to show an annual cycle. The only exceptions are the body tide, which can be modeled and removed to an accuracy far better than the accuracy of the GRACE observations, and the response of the solid Earth to the surface mass load. That loading signal, which is typically only a few percent of the signal from the load itself, can be linearly related to the load signal through scale-dependent, well-modeled, proportionality factors (load Love numbers; see succeeding text).

Thus, the seasonal mass anomaly can be assumed to be concentrated within a few kilometers of the surface. The

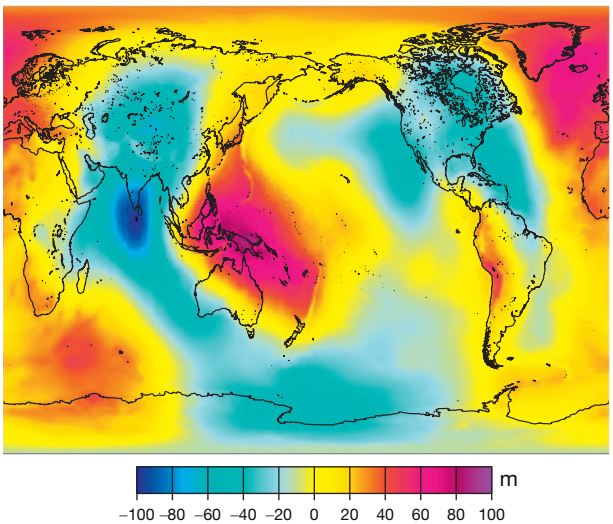


Figure 1 The time-averaged geoid anomaly from EGM96 (Lemoine et al., 1998).

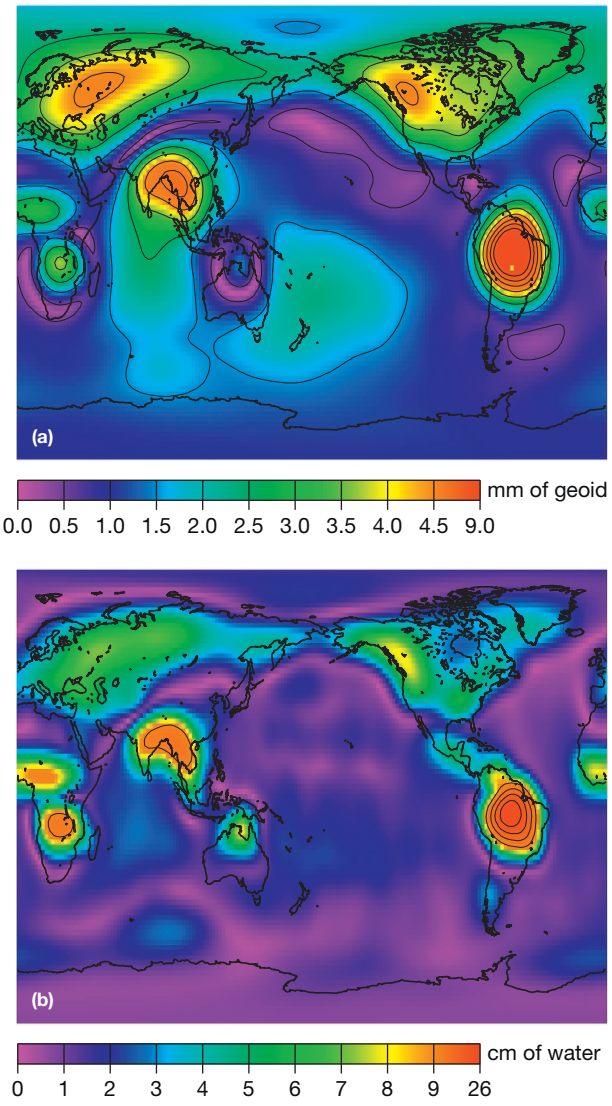


Figure 2 The top panel shows the amplitude of the annual cycle in the geoid, fit to 10 1/2 years (April 2002–September 2012) of monthly gravity field solutions from GRACE. A 750 km Gaussian smoothing function has been applied to the results. The bottom panel shows the corresponding amplitude of the annual cycle in surface mass, also smoothed with a 750 km Gaussian, in units of centimeters of water thickness. The solid lines represent contour intervals of 1 mm in the top panel and 4 cm in the bottom panel.

inversion for mass anomalies still depends, in principle, on the exact depth of the load. But since the few kilometer uncertainty in vertical position is much smaller than the horizontal scales of the signals shown in Figure 2(a), the corresponding uncertainty in the amplitude of the inferred mass anomaly is negligible. It is still not possible to tell, without additional information, whether a mass anomaly in a continental region, for example, is in the atmosphere or in the water and snow on the surface or in the water stored underground. But at least the total amplitude of the mass anomaly can be uniquely determined.

The difficulty with time-variable gravity is that the amplitudes are small. A comparison of Figures 1 and 2(a), for example, shows that the annually varying geoid is more than 1000 times smaller than the lateral variation in the static field. Most of the Earth's mass, after all, is in its rocky interior and remains relatively immobile on human timescales.

Advances in ground-based instrumentation over the past few decades have made it possible to observe time-variable gravity at local scales. Modern, high-precision gravimeters can detect surface displacements caused by solid Earth processes and local gravitational changes caused by mass variations not only in the solid Earth but also in the overlying atmosphere and underlying soil moisture and groundwater.

But the recovery of time-varying signals at large scales requires satellite measurements. Until the launch of CHAMP (Challenging Microsatellite Payload) in 2001 and, especially, GRACE in 2002, satellite time-variable gravity solutions were based entirely on satellite laser ranging (SLR) observations. The most useful SLR measurements have involved LAGEOS (launched by NASA in 1976) and LAGEOS II (launched jointly by NASA and the Italian Space Agency in 1993). Both satellites are orbiting at a 6000 km altitude and are still providing data. They are passives spheres, with outer surfaces covered with corner-cube reflectors. A powerful laser on the Earth sends a laser pulse up to the satellite, where the light is reflected back to the laser. The round-trip travel time is measured, and so the distance between the laser and the satellite is determined. By monitoring these distances from lasers around the Earth's surface, the satellite's orbital motion is computed. Since the orbital motion is determined by the Earth's gravity field, this allows for global gravity field solutions at regular time intervals. Differences between solutions for different time periods provide estimates of time-variable gravity.

3.08.1.3 Changes in the Earth's Oblateness

The first satellite identification of a nontidal time-varying signal was the recovery of a secular change in the Earth's oblateness. The oblateness is a global-scale component and is the easiest laterally varying component to detect with a satellite. There are two reasons for this. Let $N(\theta, \phi)$ be the height of the geoid above the Earth's mean spherical surface at colatitude θ and eastward longitude ϕ. It is usual to expand N as a sum of the Legendre functions (see, e.g., Chao and Gross, 1987):

$$N(\theta, \phi) = a\sum_{l=2}^{\infty}\sum_{m=0}^{l}\widetilde{P}_{lm}(\cos\theta)(C_{lm}\cos(m\phi) + S_{lm}\sin(m\phi)) \quad [1]$$

where a is the radius of the Earth, $\widetilde{P}_{lm}$ are normalized associated Legendre functions, and the C_{lm} and S_{lm} are dimensionless (Stokes) coefficients. Global gravity field solutions are typically provided in the form of a set of Stokes coefficients. The indices l and m in [1] are the degree and order, respectively, of the Legendre function. The horizontal scale of any term in [1] is inversely proportional to the value of l. The half-wavelength of a (l, m) harmonic, $(20\,000/l)$ km, traditionally serves as a representation of this scale, although we show later in the text that for the most common GRACE analyses methods, a better representation is $\approx(8500/l)$ km. Note that the sum over l in [1] begins at $l=2$. The $l=0$ term vanishes because N is defined as the departure from the mean spherical surface; and the $l=1$ terms vanish if the geoid is defined to be centered about the Earth's center of mass. Thus, the $l=2$ terms are the longest wavelength terms in the series expansion [1]. The Earth's oblateness is proportional to C_{20}.

Satellite determinations of gravity are sensitive to the gravity field at the altitude of the satellite, not at the Earth's surface. And the gravitational potential from any (l, m) term in [1] decreases with increasing radius, r, as $(a/r)^{(l+1)}$. Thus, terms with the smallest values of l (i.e., the longest wavelengths) are the least attenuated up at the satellite altitude and so tend to be the easiest to determine. This tends to favor the recovery of $l=2$ Stokes coefficients, relative to coefficients with $l>2$.

At the same time, terms with $m=0$ are better determined than terms with $m>0$. This is because an $m=0$ term does not depend on longitude. For example, Figure 3 shows the patterns of $(l, m)=(2, 0)$ and $=(2, 2)$ terms. Suppose you track a

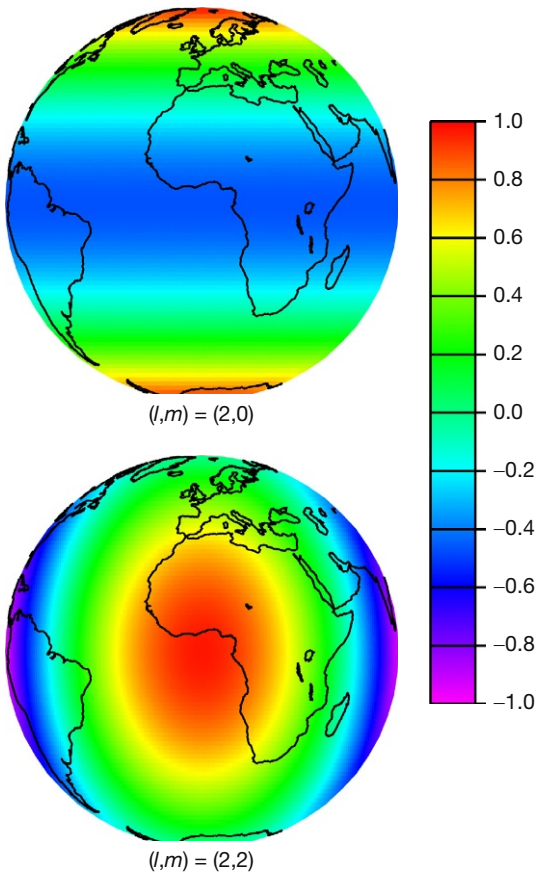


Figure 3 The spatial patterns of the Legendre functions $\widetilde{P}_{20}$ and $\widetilde{P}_{22}$.

satellite orbiting in the (2, 0) pattern shown in panel (a). As the satellite makes its first orbit, traveling from near the North Pole down to near the South Pole and back again, it passes through the gravity pattern of red/green/blue shown in the figure, and its orbit gets perturbed accordingly. By the time it begins its second orbit, the Earth has rotated about the polar axis, but because there is no longitude dependence, the satellite passes through the same red/green/blue pattern on its second orbit, and so that orbit gets perturbed in the same direction. This happens for every orbit, so the perturbation gradually builds up to large values and is easily seen in the ranging observations. On the other hand, for the (2, 2) pattern in panel (b), every time the satellite begins a new orbit, the underlying pattern is different because the Earth's rotation has carried that pattern to the east. Thus, the orbital perturbations do not tend to add constructively and are harder to see.

Early SLR solutions showed a secular increase in C_{20} (Rubincam, 1984; Yoder et al., 1983), which is consistent with a steady migration of mass from low latitudes toward high latitudes. The signal was first interpreted as due to post-glacial rebound (PGR), the Earth's ongoing response to the removal of the ice loads at the end of the last ice age. The areas that lay beneath the ice loads centered over Hudson Bay and over the region around the North and Baltic Seas are still depressed from the weight of those ancient ice sheets, and they are still gradually uplifting as material deep within the mantle flows in from lower latitudes. In fact, since its first detection, the observed secular change in C_{20} has been used in PGR models to help constrain the Earth's viscosity profile.

Figure 4 shows a recent SLR C_{20} time series (Cheng et al., 2013). There is a large seasonal variability, due presumably to a combination of atmospheric pressure variations and variations in the distribution of water in the oceans and on land (e.g., Chao and Au, 1991; Cheng and Tapley, 1999; Dong et al., 1996; Lavallee et al., 2010; Nerem et al., 2000). Long-period variability is also clearly evident, especially in the low-pass filtered version of the data (the orange line in **Figure 4**). There are interannual variations, such as the anomalous wiggle during 1998–2002, which have been variously explained as the result of climatically driven oscillations in the ocean (Cox and Chao, 2002; Dickey et al., 2002); in the storage of water, snow, and ice on land (Dickey et al., 2002); and as partly the consequence of the effects of anelasticity on the 18.6-year solid Earth tide (Benjamin et al., 2006). But there is also a dramatic and sustained change in slope beginning in the mid-1990s that has been hypothesized as the result of increasing Greenland and Antarctic ice loss rates and that has been used to conclude that those increased rates probably began about then (Nerem and Wahr, 2011).

Neither that hypothesis nor the conclusions about the sources of the interannual and seasonal variability can be inferred directly from the C_{20} data. The knowledge of just a single harmonic, like C_{20}, is not sufficient to determine the spatial location of a signal. SLR has provided time-variable solutions for a handful of other harmonics (Cheng and Tapley, 1999; Cheng et al., 1997; Moore et al., 2005; Nerem et al., 2000). But there are not nearly enough of these harmonics to give the spatial resolution necessary to confidently address these issues. The basic limitation comes from the high altitudes of LAGEOS (6000 km) and the other SLR satellites. Shorter-scale harmonics in [1] are sufficiently attenuated at those high altitudes that their time dependence cannot be easily detected. The solution to this problem is to use a satellite in a lower-altitude orbit. That was the motivation for CHAMP (Reigber et al., 2002) and, especially, for GRACE (Tapley et al., 2004a,b). In fact, the hypothesis that the change in the C_{20} slope was due to Greenland and Antarctica came from comparing the SLR values with GRACE Greenland and Antarctic estimates.

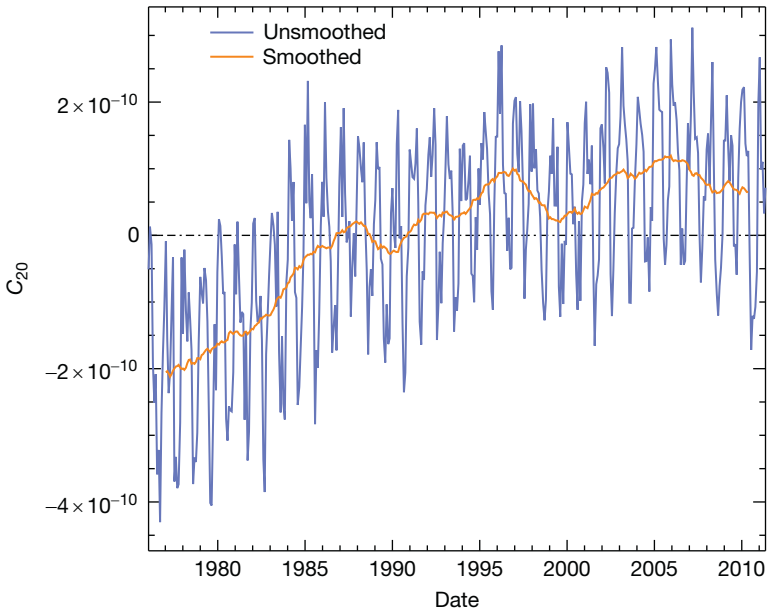


Figure 4 The blue line shows a time series of C_{20}, determined from more than 35 years of SLR measurements (Cheng et al., 2013; data provided by Minkeng Cheng). The orange line is a smoothed version of those values.

3.08.2 GRACE

The GRACE mission design makes it particularly useful for time-variable gravity studies. Launched jointly by NASA and the German Space Agency (DLR) in March 2002, GRACE consists of two identical satellites in identical orbits, one following the other by about 220 km. The satellites use microwaves to continually monitor their separation distance to an accuracy of better than 1 μm – about 1/100th the thickness of a human hair. This distance changes with time as the satellites fly through spatial gradients in the gravity field, and so by monitoring those changes, the gravity field can be determined. The satellite altitude is less than 500 km, which makes GRACE considerably more sensitive than SLR to short wavelength terms in the gravity field. The disadvantage of such a low altitude is that GRACE experiences greater atmospheric drag, which can cause large and unpredictable changes in the inter-satellite distance. To reduce this problem, each GRACE satellite has an on-board accelerometer to measure nongravitational accelerations. Those measurements are transmitted to the ground where they are used to correct the satellite-to-satellite distance measurements. Each spacecraft also has an on-board Global Positioning System (GPS) receiver, used to determine the orbital motion of each spacecraft in the GPS reference frame and so to improve the gravity field solutions at global-scale wavelengths.

3.08.2.1 Gravity Solutions

GRACE transmits raw science instrument and satellite housekeeping data to the ground, where they are transformed into physically meaningful quantities: for example, satellite-to-satellite distances, nongravitational accelerations, and spacecraft attitudes. These quantities, called level 1 data, are made publically available and can be used to construct gravity field solutions. Since few users have the capability of constructing their own gravity solutions from these data, the GRACE Project does that as well and makes those solutions, referred to as level 2 data, available on the web.

The level 2 gravity products consist of complete sets of harmonic (Stokes) coefficients [1] out to some maximum degree and order (typically $l_{max}=60$–90), averaged over monthly intervals. Larger sets of coefficients averaged over longer time intervals are also provided to represent the static field. Harmonic coefficients can be used to generate geoid, gravity, or mass solutions at individual locations or averaged over specific regions, as described later in the text. Level 2 products are generated at three project-related processing centers (i.e., the Center for Space Research at the University of Texas, GeoForschungsZentrum in Potsdam, Germany, and the Jet Propulsion Laboratory), and each of these products is made available to users. There are several independent centers around the globe that generate and distribute their own harmonic solutions, sometimes making different choices for time resolution and maximum degree.

Harmonic solutions are traditional in satellite geodesy. Harmonics help with the problem of upward and downward continuing the gravity field between the surface and the satellite altitude during the solution process. Specifically, the gravitational potential caused by any (l, m) term in [1] has a particularly simple radial dependence, decreasing with increasing radius, r, as $(a/r)^{(l+1)}$.

Nevertheless, users sometimes generate nonharmonic solutions directly from the level 1 data. Various methods have been derived for doing this, most of which involve using the satellite-to-satellite distance measurements provided in the level 1 data to directly solve for regional gravity field signals (see, e.g., Han et al., 2005a; Jekeli, 1999; Schmidt et al., 2006a).

Another approach involves the construction of 'mascon' solutions directly from the level 1 data (e.g., Rowlands et al., 2005). Mascons, in this context, are mass anomalies spread uniformly over either regular-shaped (usually rectangular) or irregular-shaped blocks (Luthcke et al., 2006) at the Earth's surface. Each such mass anomaly has an overall scale factor, which is determined by fitting to the level 1 satellite-to-satellite data.

3.08.2.2 Using the Harmonic Solutions to Solve for Mass

Most users do not have the resources to process level 1 data and rely instead on the standard level 2 gravity field products: the harmonic solutions. For most applications, the gravity field itself is not of direct interest. Instead, it is usually the mass distribution causing the gravity field that is the desired quantity. Here, we describe how that mass distribution can be inferred from the harmonic gravity solutions. We focus specifically on the time-variable components of the gravity and mass fields. The methods described here are described in more detail in **Section 3.08.2.1** of Wahr et al. (1998) (see also Chao and Gross, 1987).

The time-variable component of the gravity field is obtained by removing the long-term mean of the Stokes coefficients from each monthly value. The mean can be taken to be one of the static fields available as level 2 products. Or, perhaps more usefully, it can be estimated by simply constructing the average of all the monthly fields used in the analysis. The reason for removing the mean field is that it is dominated by the static density distribution inside the solid Earth. It thus has no bearing on attempts to learn about, say, the distribution of water stored on land or in the ocean. Removing the static field, though, means that all contributions from the time-averaged value of the stored water are also removed. Thus, only the time-variable component of the water storage can be recovered.

The time-variable gravity field is then used to solve for the time-variable mass field. This solution is nonunique, as described in the 'Introduction.' Let ΔC_{lm} and ΔS_{lm} be the time-variable components of the (l, m) Stokes coefficients for some month. Let $\Delta\rho(r, \theta, \phi)$ be the density redistribution that causes this time-dependent change in gravity. Then,

$$\begin{Bmatrix} \Delta C_{lm} \\ \Delta S_{lm} \end{Bmatrix} = \frac{3}{4\pi a \rho_{ave}(2l+1)} \int \Delta\rho(r,\theta,\phi) \left(\frac{r}{a}\right)^{l+2} \widetilde{P}_{lm} (\cos\theta) \begin{Bmatrix} \cos(m\phi) \\ \sin(m\phi) \end{Bmatrix} \sin\theta \, d\theta \, d\phi \, dr \qquad [2]$$

where ρ_{ave} is the Earth's average density ($=5517$ kg m^{-3}).

Suppose the density is expanded as a sum of the Legendre functions:

$$\Delta\rho(r,\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=0}^{l}\widetilde{P}_{lm}(\cos\theta)\left(\Delta\rho^{c}_{lm}(r)\cos(m\phi)+\Delta\rho^{s}_{lm}(r)\sin(m\phi)\right) \quad [3]$$

Using [3] in [2] and employing orthogonality relations for Legendre functions, [2] reduces to

$$\begin{Bmatrix}\Delta C_{lm}\\ \Delta S_{lm}\end{Bmatrix}=\frac{3}{a\rho_{\text{ave}}(2l+1)}\int\begin{Bmatrix}\Delta\rho^{c}_{lm}(r)\\ \Delta\rho^{s}_{lm}(r)\end{Bmatrix}\left(\frac{r}{a}\right)^{l+2}\mathrm{d}r \quad [4]$$

This result [4] can be used to place constraints on $\Delta\rho(r, \theta, \phi)$ from measurements of ΔC_{lm} and ΔS_{lm}. The nonuniqueness is evident here in the fact that ΔC_{lm} and ΔS_{lm} provide information only on the radially weighted radial integral of the density coefficients. There is no way of determining how the density depends on depth within the Earth.

Suppose, though, we have reason to believe the observed ΔC_{lm} and ΔS_{lm} are caused by mass variability concentrated within a thin layer of thickness H near the Earth's surface, a layer containing those regions of the atmosphere, oceans, ice sheets, and land water storage that are subject to significant mass fluctuations. H, in this case, would be mostly determined by the thickness of the atmosphere and is of the order of 10 km or less. If H is thin compared to the horizontal resolution of the observations, then the amplitude of the density anomaly can be uniquely determined, as follows.

Suppose the observed gravity field is accurate enough to resolve gravity anomalies down to scales of R km. That means the ΔC_{lm} and ΔS_{lm} contain useful information for values of l up to $l_{\max}=(8500/R)$ km (see succeeding text). At present, the individual monthly GRACE solutions have a typical resolution of ~500 km near the equator, improving to ~250 km at high latitudes. (Even better resolutions can be obtained by employing postprocessing methods (e.g., Swenson and Wahr, 2006a).) Thus, at present, $l_{\max}\sim 35$. Suppose H is thin enough that

$$(l_{\max}+2)H/a\ll 1 \quad [5]$$

Then, since r in the integrand of [4] differs from a by no more than H, $(r/a)^{l+2}$ differs from 1 by no more than $[(a+H)/a)^{l+2}-1]\approx 1+(l+2)H/a\approx 1$ for all $l\le l_{\max}$. In this case, [2] reduces to

$$\begin{Bmatrix}\Delta C^{\text{surf mass}}_{lm}\\ \Delta S^{\text{surf mass}}_{lm}\end{Bmatrix}=\frac{3}{4\pi a\rho_{\text{ave}}(2l+1)}\int\Delta\sigma(\theta,\phi)\widetilde{P}_{lm}(\cos\theta)\begin{Bmatrix}\cos(m\phi)\\ \sin(m\phi)\end{Bmatrix}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \quad [6]$$

where $\Delta\sigma$ is the change in surface density (i.e., mass/area), defined as the radial integral of $\Delta\rho$ through the surface layer:

$$\Delta\sigma(\theta,\phi)=\int_{\text{thin layer}}\Delta\rho(r,\theta,\phi)\mathrm{d}r \quad [7]$$

The assumption that the density anomaly is concentrated entirely within this thin layer is incorrect. Any change in mass load at the surface will induce deformation within the solid Earth, leading to a density anomaly at depth as well. The gravity signal caused by these solid Earth mass anomalies is typically a few percent of the gravity anomaly caused by the surface mass and fortunately can be easily represented in terms of load Love numbers, k_l (see, e.g., Chao, 1994; Farrell, 1972; eqn [6]). Specifically, if $\Delta C^{\text{solidEarth}}_{lm}$ and $\Delta S^{\text{solidEarth}}_{lm}$ represent the contributions to the gravity field from the load-induced deformation in the solid Earth, then

$$\begin{Bmatrix}\Delta C^{\text{solid Earth}}_{lm}\\ \Delta S^{\text{solid Earth}}_{lm}\end{Bmatrix}=k_l\begin{Bmatrix}\Delta C^{\text{surf mass}}_{lm}\\ \Delta S^{\text{surf mass}}_{lm}\end{Bmatrix} \quad [8]$$

Thus, the total dependence of the Stokes coefficients on the surface mass density is

$$\begin{Bmatrix}\Delta C_{lm}\\ \Delta S_{lm}\end{Bmatrix}=\frac{3}{4\pi a\rho_{\text{ave}}}\frac{1+k_l}{(2l+1)}\int\Delta\sigma(\theta,\phi)\widetilde{P}_{lm}(\cos\theta)\begin{Bmatrix}\cos(m\phi)\\ \sin(m\phi)\end{Bmatrix}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \quad [9]$$

By expanding $\Delta\sigma(\theta, \phi)$ as a sum of Legendre coefficients, similar to the expansion shown in [3] for $\Delta\rho$, and using the orthogonality of the Legendre functions to obtain a result similar to [4], we find

$$\Delta\sigma(\theta,\phi)=\frac{a\rho_{\text{ave}}}{3}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\frac{2l+1}{1+k_l}\widetilde{P}_{lm}(\cos\theta)(\Delta C_{lm}\cos(m\phi)+\Delta S_{lm}\sin(m\phi)) \quad [10]$$

The results in the preceding text assume the surface layer is thin enough that [5] is valid. If we assume that $l_{\max}=35$ and that the layer includes the atmosphere so that $H\sim 10$ km or less, then [5] is violated at only about the 5% level. This is a large enough inaccuracy that it might be important, for some applications, to include the radial distribution of atmospheric density fluctuations (Swenson and Wahr, 2002a). Mass variations in the oceans and in the water stored on land occur almost entirely within 1 km of the surface and usually much closer to the surface than that. For a 1 km thick layer and $l_{\max}=35$, [5] is accurate to ~0.5%, which is easily good enough for oceanographic and hydrologic applications.

3.08.2.3 Love Numbers

The use of [10] to recover surface mass requires knowledge of the load Love numbers k_l. As a guide, one set of results for those Love numbers is given in **Table 1** (D. Han, personal communication, 1998) for a few values of l up to 200. These results are computed as described by Han and Wahr (1995), using Earth structural parameters from the Preliminary reference Earth model (PREM) of Dziewonski and Anderson (1981). Results for other values of $l<200$ can be obtained by linear interpolation of **Table 1** results. Linearly interpolating **Table 1** results, instead of using exact results, introduces errors in the k_l of less than 0.05% for all $l<200$.

These results for k_l do not include anelastic effects. Those effects increase with increasing period but are apt to be negligible for our applications. For example, Wahr and Bergen (1986) concluded that at an annual period, the anelastic effects on the $l=2$ body tide Love number, k_2^{body}, would probably be less than 2%, corresponding to an effect on $(1+k_2^{\text{body}})$ of less than 1%. Even allowing for larger effects at longer periods and perhaps a somewhat greater effect for load Love numbers than for body tide Love numbers (since load Love numbers are more sensitive to upper mantle structure where the anelastic effects could be larger), we tentatively conclude that anelasticity would not perturb the results for $(1+k_l)$ by more than a few percent.

Table 1 Elastic Love numbers k_l computed by Dazhong Han as described by Han and Wahr (1995), for Earth model PREM (Dziewonski and Anderson, 1981)

l	k_l
0	+0.000
1	+0.027
2	−0.303
3	−0.194
4	−0.132
5	−0.104
6	−0.089
7	−0.081
8	−0.076
9	−0.072
10	−0.069
12	−0.064
15	−0.058
20	−0.051
30	−0.040
40	−0.033
50	−0.027
70	−0.020
100	−0.014
150	−0.010
200	−0.007

The $l=1$ value assumes that the origin of the coordinate system is the center of figure of the solid Earth's surface (see text).

The Love numbers in [10] with $l=0$ and $l=1$ require discussion. The $l=0$ term is proportional to the total mass of the Earth where 'the Earth' includes not only the solid Earth but also its fluid envelope (the oceans, atmosphere, etc.). This total mass does not change with time, and so ΔC_{00} from GRACE can be assumed to vanish. Suppose, though, the objective is to use [10] to find the surface mass contribution from just one component of the surface mass: the ocean, for example. The total mass of the ocean need not be constant, due to the exchange of water with the atmosphere or the land surface. So the oceanic contributions to $\Delta\hat{C}_{00}$ need not vanish. But this nonzero $\Delta\hat{C}_{00}$ will not induce an $l=0$ response in the solid Earth: that is, the load does not cause a change in the total solid Earth mass. Thus, $k_0=0$.

The $l=1$ terms are proportional to the position of the Earth's center of mass relative to the center of the coordinate system and so depend on how the coordinate system is chosen. One possibility is to choose a system where the origin always coincides with the Earth's instantaneous center of mass. In that case, all $l=1$ terms in the geoid are zero by definition, and so the GRACE results for $\Delta C_{lm}=\Delta S_{lm}=0$ for all $l=1$. This is the coordinate system used for the geoid representation shown in [1]. Again, the $l=1$ coefficients for an individual component of the total surface mass need not vanish. The redistribution of mass in the ocean, for example, can change the center of mass of the ocean. But that will induce a change in the center of mass of the solid Earth, so that the center of mass of the ocean + solid Earth remains fixed. So, for this choice of coordinate system, $k_{l=1}=-1$.

Another possibility is to define the coordinate system so that its origin coincides with the center of figure of the Earth's solid outer surface. That is the most sensible way of defining the origin when recovering the Earth's time-variable mass distribution, since hydrologic, oceanographic, and atmospheric models are invariably constructed in a system fixed to the Earth's surface. In that case, the $l=1$ GRACE results for $\Delta C_{lm}=\Delta S_{lm}$ need not vanish, and the Love number $k_{l=1}$ is defined so that the $l=1$ terms in [10] describe the offset between the center of mass of the surface mass + deformed solid Earth and the center of figure of the deformed solid Earth surface. It is shown by Trupin et al. (1992; eqn [10]) that for this coordinate system, $k_{l=1}=-(h_{l=1}+2\ell_{l=1})/3$, where $h_{l=1}$ and $\ell_{l=1}$ are the $l=1$ displacement Love numbers when the origin is the center of mass of the deformed solid Earth. For this choice of origin, the numerical value of $k_{l=1}=-(h_{l=1}+2\ell_{l=1})/3$ is given in **Table 1**.

3.08.2.4 Spatial Averaging

Equation [10] is the starting point for using GRACE estimates of ΔC_{lm} and ΔS_{lm} to recover changes in surface mass density. Because the errors in the GRACE results become large for large l (i.e., short scales) and because terms with large l values can make important contributions to the sum in [10] (note the $2l+1$ factor in the numerator of [10]), the use of [10] as written can lead to inaccurate results.

To obtain accurate results, it is necessary to somehow reduce the large l contributions to the sum [10]. This involves the insertion of some additional multiplicative factor into [10] that is small for large values of l. Any such modification means that the sum will no longer be an exact representation of the surface mass at (θ, ϕ). Since most applications require the surface mass in the spatial domain, it is useful to choose a multiplicative factor in such a way that the sum still has some meaningful connection to the spatially dependent surface mass. Any multiplicative factor applied in the spectral (l, m) domain is equivalent to convolving with some corresponding weighting function in the spatial domain. The problem is to choose a factor that reduces the errors but that keeps the weighting function localized. The issues are similar to those encountered when designing time series filters, where the generic problem is to construct a filter that removes noise but that still provides a meaningful estimate of the true signal in the time domain.

Various methods have been used for improving the GRACE mass estimates in this way, though most of them are similar to one another (e.g., Chen et al., 2006a; Han et al., 2005b; Seo and Wilson, 2005; Swenson and Wahr, 2002b; Wahr et al., 1998). Most of these methods fall into one of two categories: smoothing the surface mass results or averaging over specific regions.

3.08.2.4.1 Smoothing

The simplest way of modifying [10] to obtain accurate results is to introduce degree-dependent weighting factors W_l into the sum, so that

$$\overline{\Delta\sigma}(\theta,\phi)=\frac{a\rho_{\text{ave}}}{3}\sum_{l,m}\frac{2l+1}{1+k_l}W_l\widetilde{P}_{lm}(\cos\theta)[\Delta C_{lm}\cos(m\phi)+\Delta S_{lm}\sin(m\phi)] \quad [11]$$

$\overline{\Delta\sigma}$ then represents a smoothed version of the surface mass anomaly, given by

$$\overline{\Delta\sigma}(\theta,\phi)=\int \sin\theta'\,d\theta'\,d\phi'\,\Delta\sigma(\theta',\phi')W(\alpha) \qquad [12]$$

where α is the angle between (θ, ϕ) and (θ', ϕ') and $W(\alpha)$ is a smoothing function corresponding to the choice of the W_l's:

$$W(\alpha)=\frac{1}{4\pi}\sum_l \sqrt{2l+1}\,W_l\widetilde{P}_{l0}(\alpha) \qquad [13]$$

One obvious way of smoothing is simply to truncate the sum over l so that the inaccurate coefficients at large l are not included. This is equivalent to choosing $W_l=1$ for values of l less than some l_{max} and $W_l=0$ for $l\geq l_{max}$. This approach can, indeed, give accurate results for the sum if l_{max} is chosen to be small enough. The disadvantage of using this step-function weighting is that the equivalent convolution function, $W(\alpha)$, 'rings' in the spatial domain (see panels (a) and (b) in Figure 5). The results for $\overline{\Delta\sigma}$ in this case are on average not only of the true values of $\Delta\sigma$ at points close to (θ, ϕ) but also of the $\Delta\sigma$ values at points all around the globe and where the smoothing function has an oscillating sign.

This ringing can be avoided by choosing W_l to decrease smoothly with l. A convenient choice of smoothing coefficients (see, e.g., Wahr et al., 1998) is the Gaussian values developed by Jekeli (1981) to improve estimates of the Earth's gravity field. Those coefficients can be found using the recursion relations

$$\begin{aligned} W_0&=1\\ W_1&=\frac{1+e^{-2b}}{1-e^{-2b}}-\frac{1}{b}\\ W_{l+1}&=-\frac{2l+1}{b}W_l+W_{l-1}\end{aligned} \qquad [14]$$

These coefficients correspond to a smoothing function

$$W(\alpha)=\frac{b\exp[-b(1-\cos\alpha)]}{1-e^{-2b}} \qquad [15]$$

where

$$b=\frac{\ln(2)}{(1-\cos(r/a))} \qquad [16]$$

and r is the distance on the Earth's surface at which W has decreased to 1/2 its value at $\alpha=0$ (the distance on the Earth's surface $=a\alpha$). We will refer to r as the smoothing radius. As an example, panels (c) and (d) of Figure 5 show W_l and $W(\alpha)$ for

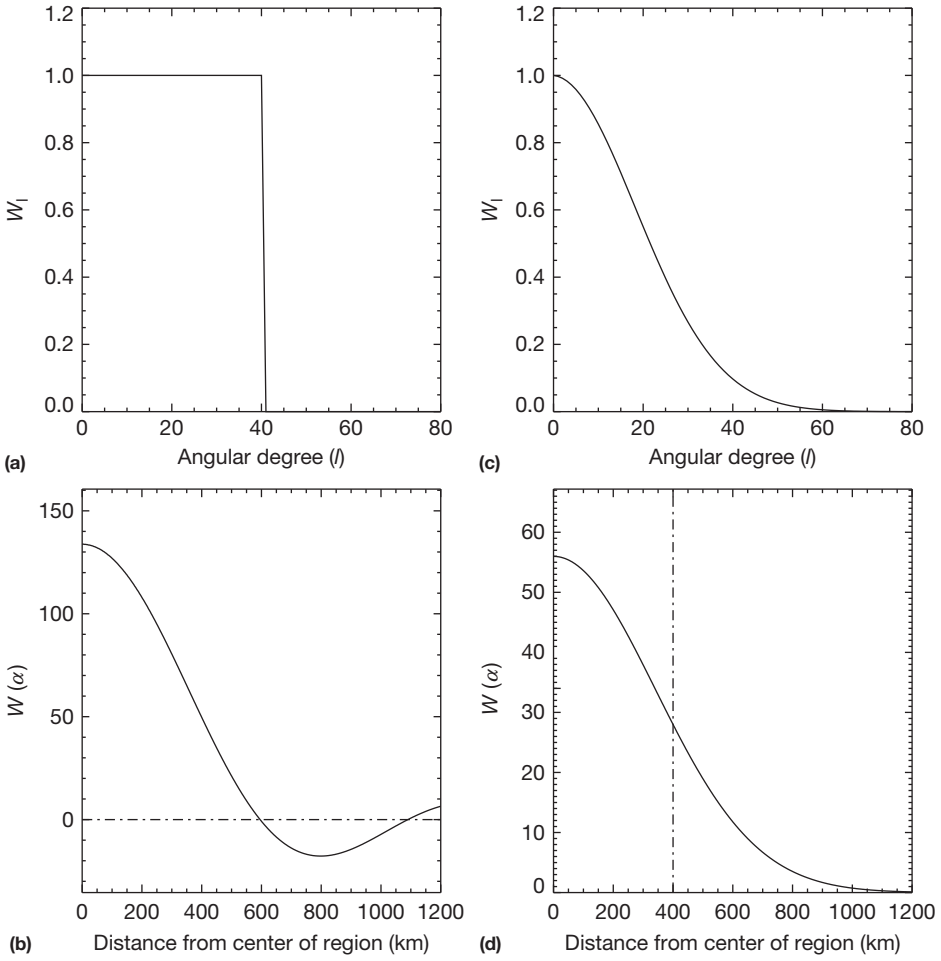


Figure 5 (a) The spectral smoothing coefficients equivalent to truncation at degree $l=40$ and (b) the corresponding smoothing function in the spatial domain. (c) The spectral smoothing coefficients for Gaussian smoothing with a 400 km radius and (d) the corresponding smoothing function in the spatial domain.

$r = 400$ km. Note that the convolution function, $W(\alpha)$, decreases smoothly to zero at large angular distances and does not oscillate. In practice, there will always be some oscillation, since no satellite gravity field model will ever provide the Stokes coefficients out to infinite degree. But as long as the W_l are small out at the value of the maximum degree in the gravity model, the ringing is minimal.

Note, also, from Figure 5(c) that at angular degree $l \approx 21$, the W_l coefficients have decreased to about half their value at $l = 0$. In general, if R_{half} is the half-width of a Gaussian filter $W(\alpha)$ and l_{half} is the half-width of the corresponding harmonic coefficients W_l, then $R_{\text{half}} \times l_{\text{half}} \approx 8500$ km. A familiar rule of thumb in geodesy is that a spherical harmonic of degree l has a spatial resolution of 20 000 km/l. But for Gaussian averaging, which is used extensively in GRACE analyses, a more appropriate relation between harmonic degree and resolution is $R_{\text{half}} \approx (8500/l_{\text{half}})$ km.

The annual amplitude in surface mass shown in Figure 2(b) is obtained by applying a Gaussian smoothing function with a 750 km radius to monthly CSR Release 5 GRACE mass solutions between April 2002 and September 2012. For comparison, the top panel of Figure 6 shows the results for a single month (July 2007; after the temporal mean has been subtracted) for a 250 km radius. Note that the significant noise that is present when this shorter averaging radius is used. This occurs because the high-degree terms in [10] are not attenuated as effectively for shorter smoothing radii. The disadvantage of using larger smoothing radii is that the results in the spatial domain are then less able to pick up short-scale structure in the mass anomalies.

The results shown in the top panel of Figure 6 suggest that 250 km resolution is beyond the current capabilities of GRACE at low- to midlatitudes, at least for single-month solutions. (The results for seasonal terms or secular trends are not as

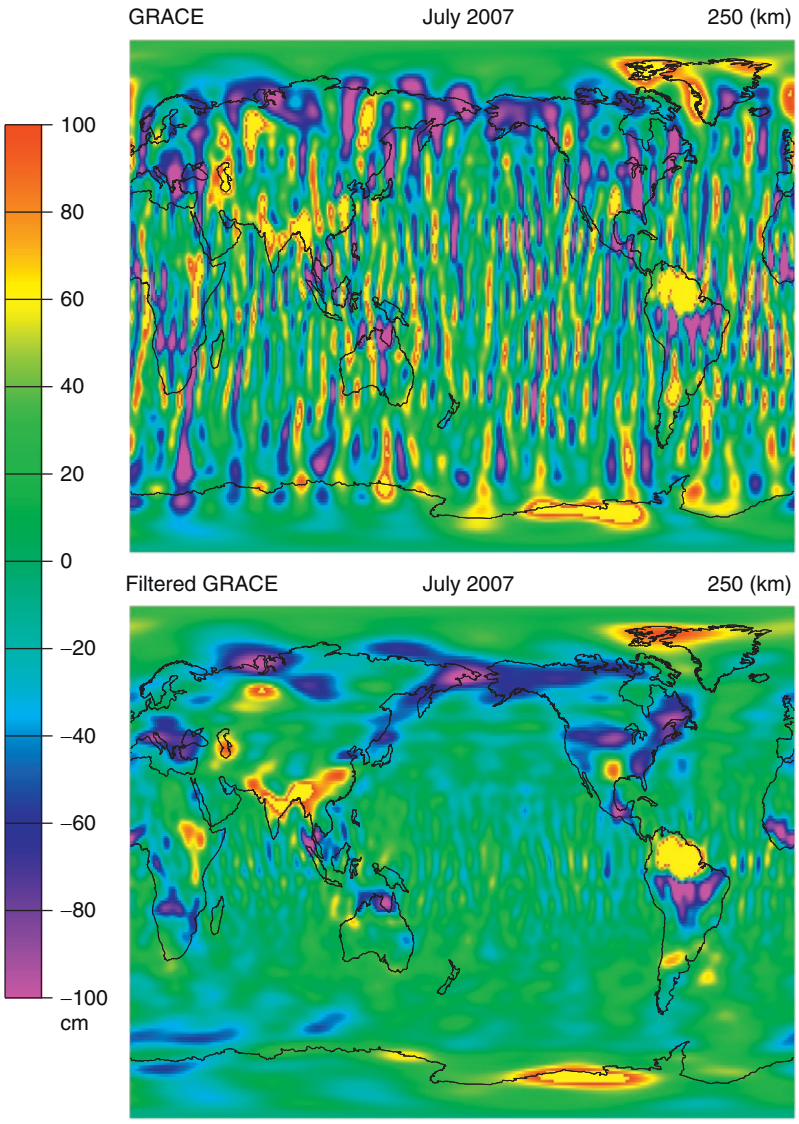


Figure 6 The top panel shows surface mass anomalies deduced from GRACE for a single month (July 2007), after the temporal mean has been removed and after smoothing with a 250 km Gaussian. The units are millimeters of water thickness. The bottom panel shows the same thing, but after postprocessing, the Stokes coefficients to reduce noise as described by Swenson and Wahr (2006a). Figure provided by Sean Swenson.

noisy, because the errors in individual months tend to be uncorrelated and so become relatively less important when fitting to multiple months.) Note that the noise seems to be oriented in north–south stripes. This stripiness of the errors is a characteristic feature of GRACE gravity solutions and is not found, for example, in SLR gravity fields. It occurs because the GRACE satellites measure gravity gradients along-track, and since the GRACE inclination is 89°, the tracks are oriented north–south. Thus, there is little east–west sensitivity and so errors in the measurements or in the processing tend to be put into east–west gradients. The reduced errors at very high latitudes occur because the satellites pass over the poles during every orbit, and so they monitor high latitudes much more thoroughly than lower latitudes.

Postprocessing methods can be used to remove the stripes. The bottom panel of **Figure 6** shows results for the same 250 km smoothing radius as the top panel, but after applying the Swenson and Wahr (2006a) correlated filter to the Stokes coefficients. Simulations show that this filter reduces stripes without seriously impacting the recovery of most real geophysical signal. Note that the stripes in the bottom panel are, indeed, greatly reduced and that features that look like true signals are now clearly evident. Still, no filtering technique will leave the true signal completely untouched. Signals that tend to appear as north–south trending features are especially susceptible to being absorbed by this particular filter. Furthermore, this filter tends to spread signals out in the east–west direction, as is evident from looking at some of the mid- to high-latitude features in the bottom panel of **Figure 6**. Thus, when employing this or any other filter – or, for that matter, using the smoothing methods described in the previous section – care must be taken into account for signal absorption and distortion when comparing with independent data or models. This is commonly done by applying a scaling factor to the result (see, e.g., Landerer and Swenson, 2012).

3.08.2.4.2 Regional averaging

Many applications require estimates of mass variability for specific regions, for example, estimating changes in mass of the Antarctic ice sheet or changes in water storage in the Mississippi River basin. These sorts of problems are better addressed by constructing specific averaging functions optimized for those regions, than by employing the sort of generic smoothing functions described in the preceding text.

For example, an exact regional average would take the form

$$\Delta\sigma_{\text{region}} = \frac{1}{\Omega_{\text{region}}}\int \Delta\sigma(\theta,\phi)\vartheta(\theta,\phi)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \qquad [17]$$

where Ω_{region} is the angular area of the region of interest and where

$$\vartheta(\theta,\phi) = \begin{cases} 0 & \text{outside the basin} \\ 1 & \text{inside the basin} \end{cases} \qquad [18]$$

The result [17] can be expressed as a sum of Stokes coefficients:

$$\Delta\sigma_{\text{region}} = \frac{a\,\rho_{\text{ave}}}{3\Omega_{\text{region}}}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\frac{(2l+1)}{(1+k_l)}\left(\vartheta^c_{lm}\Delta C_{lm} + \vartheta^s_{lm}\Delta S_{lm}\right) \qquad [19]$$

where ϑ^c_{lm} and ϑ^s_{lm} are the harmonic coefficients of $\vartheta(\theta,\phi)$. Since the averaging function, $\vartheta(\theta,\phi)$ in this case, changes abruptly from 1 to 0 along the edge of the region, it has power at short spatial scales. Thus, ϑ^c_{lm} and ϑ^s_{lm} can be relatively large at high degrees, and so this estimate of $\Delta\sigma_{\text{region}}$ can be inaccurate.

A way around this problem is to smooth the averaging function, so that it is close to 1 inside the region and close to 0 outside and varies smoothly between 0 and 1 along the edges. We replace [17] with

$$\overline{\Delta\sigma}_{\text{region}} = \frac{1}{\Omega_{\text{region}}}\int \Delta\sigma(\theta,\phi)\overline{W}(\theta,\phi)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \qquad [20]$$

where the averaging function $\overline{W}(\theta,\phi)$, which can be expanded as

$$\overline{W}(\theta,\phi) = \frac{1}{4\pi}\sum_{lm}\widetilde{P}_{lm}(\cos\theta)\left\{W^c_{lm}\cos(m\phi) + W^s_{lm}\sin(m\phi)\right\} \qquad [21]$$

is chosen to closely approximate $\vartheta(\theta,\phi)$, but to vary smoothly enough that its expansion coefficients W^c_{lm} and W^s_{lm} are small for large values of l. In that case, the spectral equivalent to [20],

$$\overline{\Delta\sigma}_{\text{region}} = \sum_{l,m}\frac{a\rho_{\text{ave}}}{3\Omega_{\text{region}}}\frac{(2l+1)}{(1+k_l)}\left(W^c_{lm}\Delta C_{lm} + W^s_{lm}\Delta S_{lm}\right) \qquad [22]$$

will be both reasonably representative of the true regional average and reasonably accurate. The methods of optimizing the choice of $\overline{W}(\theta,\phi)$, based on estimates of the true signal characteristics, are described by Swenson and Wahr (2002b) and Seo and Wilson (2005). In general, the larger the region, the more accurate the results are. The examples of optimal averaging functions for the Mississippi River basin and for Antarctica are shown in **Figure 7**. Note that in both cases, the averaging function is smaller than 1 inside the region (especially evident for the Mississippi River basin) and remains larger than 0 for some distance outside the region.

3.08.2.5 Estimating Errors and Accounting for Leakage

Errors in a surface mass estimate separate into two categories: those due to errors in the Stokes coefficients and those caused by leakage from other signals. Errors in the monthly averaged Stokes coefficients can be caused by instrumental errors, data processing problems, or unmodeled temporal aliasing signals. Temporal aliasing errors in the GRACE monthly gravity fields are caused by short-period (submonthly) variations in gravity. The satellite does not monitor the entire global field continuously during a month, but samples the gravity field only along its orbital path. Infrequent sampling of a short-period signal can cause aliasing into the monthly averages. The best way to reduce these aliasing errors is to independently model and remove the effects of short-period gravity variations before constructing monthly averages. For GRACE, this means modeling and removing the effects of solid Earth and ocean tides, of atmospheric mass variability over land (using global, gridded atmospheric fields available from the European Centre for Medium-Range Weather Forecasts), and of short-period variations in ocean bottom pressure (using an ocean general circulation model). Errors in any of those models cause aliasing errors in the monthly gravity field solutions (Han et al., 2004, 2005c;

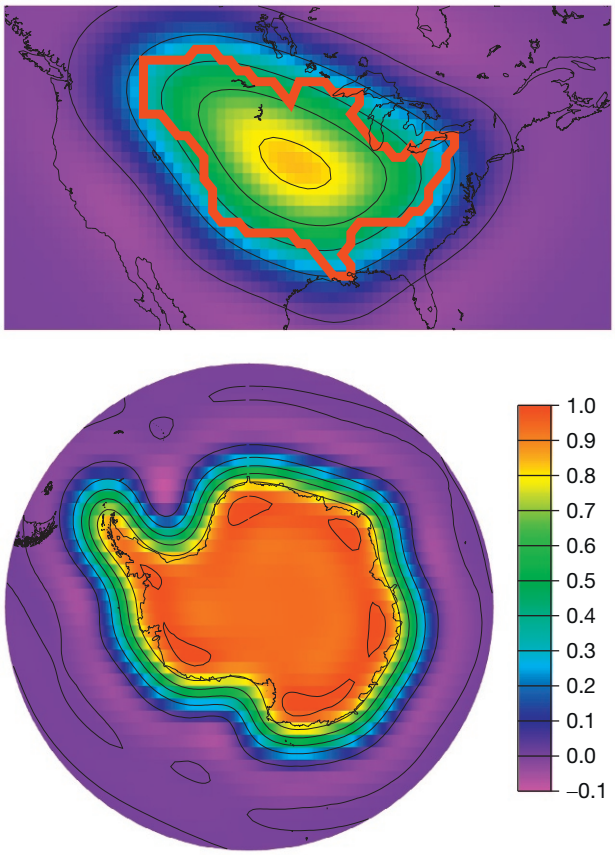


Figure 7 Averaging functions, $\overline{W}$ (see eqn [21]), for the Mississippi River basin and for Antarctica. The solid lines represent contour intervals of 0.2.

Knudsen and Andersen, 2002; Ray and Luthcke, 2006; Schrama, 2004; Song and Zlotnicki, 2004; Thompson et al., 2004).

To see how errors in the Stokes coefficients from any source (i.e., instrumental, processing, or aliasing) map into errors in a mass estimate, let δC_{lm} and δS_{lm} be the root-mean-square (rms) errors in the Stokes coefficients. The smoothed estimates [11] and the regional averages [22] are both of the form

$$\sigma = \sum_{l,m} \left[F_l^m \Delta C_{lm} + G_l^m \Delta S_{lm} \right] \qquad [23]$$

Suppose the errors in the different Stokes coefficients are uncorrelated with one another. Then, the corresponding rms error in σ would be (Wahr et al., 2006)

$$\delta\sigma = \sqrt{\sum_{l,m} \left(F_{lm}^2 \delta C_{lm}^2 + G_{lm}^2 \delta S_{lm}^2 \right)} \qquad [24]$$

The errors in different Stokes coefficients are likely to be correlated. For GRACE, those correlations are responsible for the stripes evident in the top panel of **Figure 6**. The knowledge of the full error covariance matrix can improve the estimate of $\delta\sigma$. But even without the full covariance, [24] provides a reasonable first approximation for $\delta\sigma$, if δC_{lm} and δS_{lm} can be estimated.

To understand leakage errors, consider an application where the goal is to use the Stokes coefficients to assess a regional water storage model. For example, suppose a surface mass average of the form [22] is constructed and interpreted as an estimate of water storage variability in some chosen river basin. Leakage errors are the contributions to [22] caused by gravity signals from outside the basin.

These leakage errors can come from time-variable mass anomalies either vertically above or below the river basin or from mass anomalies off to the side of the basin. Signals above or below would come from the overlying atmosphere or the underlying solid Earth and cannot be separated from the river basin signal no matter how complete and accurate the gravity field estimation. This is a consequence of the nonuniqueness of gravity-based inversions for density, as described in the preceding text. The only recourse is to independently model and remove the atmospheric and solid Earth signals. Any inaccuracy in those models is thus a source of errors for the hydrology estimates (Velicogna et al., 2001).

Leakage from mass anomalies off to the side, in neighboring river basins, for example, can be minimized using a weighting function that is as localized as possible to the river basin of interest. As described previously, though, an averaging function should usually be constructed to be smoother than the basin function, to provide an accurate estimate.

For some applications, this horizontal leakage is not an issue. For example, suppose the objective is to compare the satellite estimates of σ for the Mississippi River basin, with the output of a hydrology model. The leakage into the satellite estimate will come mainly from the river basins that border the Mississippi River basin. If the same averaging function is applied to the model output, then both will be subject to the same leakage. The model–satellite comparison will then actually be a comparison over a somewhat broader region than just the Mississippi River basin, but the model and satellite results will both be affected by leakage in the same way.

But for many applications, the goal is to estimate mass variability within a specific region with no contamination from outside. In that case, leakage is an inescapable source of error. The best way to estimate the likely impact of that error is to apply the averaging function to simulated data. This sort of problem commonly arises in time series analysis, where our averaging process is analogous to low-pass filtering. A low-pass filter not only removes high frequencies but also reduces the low-frequency signal; that is, each filter has a characteristic gain function. The effects of the gain function must be determined and removed from the filtered data, to estimate the true low-frequency signal in the time domain.

The examples shown in **Figure 7**, that is, the Mississippi River basin and Antarctica, illustrate two types of situations. For the Mississippi, the averaging function downweighs the true Mississippi signal, since the averaging function is smaller than 1 everywhere within the basin. In effect, the averaging function replaces some of the signal located inside the Mississippi River basin, with signals located outside in neighboring basins. The effects of this leakage depend on whether the external hydrology signal looks like the internal signal and so basically comes down to a comparison between the correlation length of the hydrology signal (which tends to be controlled by the scale length of the precipitation) and the resolution of the averaging kernel (which is usually based on the resolution of the gravity field). For a reasonably homogeneous region like

this portion of the interior of the United States, the signal just outside the basin is similar enough to the signal just inside, in that the leakage from the averaging kernel shown in Figure 7 might not be too severe. Nonetheless, the leakage is nonzero and should be estimated using hydrology model output (Landerer and Swenson, 2012).

For Antarctica, the extension of the averaging function over the ocean means that some Antarctic signal is being replaced by ocean signal. There is likely to be no correlation at all between the Antarctic and ocean signals. For example, suppose the goal is to determine the linear trend in Antarctic mass over some multiyear period. It is probable that there would be little or no multiyear trend over the ocean. So the averaging process underrepresents the contribution from the trend in Antarctic mass and replaces it with a negligible trend from the ocean. This leads to an underestimate of the Antarctic mass trend. The situation is similar for any region where the signal of interest is much larger than the signal in surrounding areas. Again, the only way to assess and correct for this effect is to apply the averaging function to simulated data for the Antarctic ice sheet and the surrounding ocean (e.g., see Velicogna and Wahr, 2006a).

In principle, the sum in [22] should include all l in the range $0 \le l \le \infty$. In practice, the sum for GRACE is limited to $2 \le l \le l_{max}$, where l_{max} can be no larger than the maximum degree of the GRACE fields. The truncation to $l \le l_{max}$ causes ringing: sensitivity to mass variability well outside the region of interest, though this sensitivity is weak if l_{max} is large. The restriction to $l \ge 2$ arises because GRACE does not recover $l = 0$ or $l = 1$ coefficients. The $l = 0$ coefficient is proportional to the Earth's total mass. Since that mass remains constant, $\Delta C_{00} = 0$ is a correct assumption. But the omission of $l = 1$ terms in [22] can cause errors in estimates of $\overline{\Delta\sigma}_{region}$. Those terms are proportional to the displacement of the geocenter (the offset between the Earth's center of mass and the center of figure of the surface) and are particularly affected by seasonal or long-term transfer of water, snow, and ice between the continents and ocean. Those terms are, indeed, zero in a coordinate system that is always centered on the Earth's center of mass. But we, along with the oceans, the ice sheets, the atmosphere, etc., all operate in a coordinate system attached to the Earth's surface, and the $l = 1$ do not vanish in that coordinate system. Their omission from [22] means, in effect, that the averaging function has a small-amplitude tail that extends around the globe, causing distant signals to leak into $\overline{\Delta\sigma}_{region}$. This leakage can be removed by augmenting the GRACE-derived gravity fields with independent estimates of time-variable geocenter coefficients obtained from other techniques (i.e., SLR or GPS). Surprisingly, geocenter values can be indirectly inferred from the GRACE $l > 1$ Stokes coefficients (Swenson et al., 2008), and these values are probably the ones used most often in GRACE analyses.

3.08.2.5.1 Other methods

Other analysis methods, besides smoothing or convolving with an averaging kernel, have sometimes been applied to the GRACE Stokes coefficients to improve the solutions. These methods include the application of optimally designed spectral windows to better focus the sensitivity kernel (Han and Simons, 2008) and the use of empirical orthogonal functions to reduce noise (Wouters and Schrama, 2007). One effective and easily implemented analysis technique is to fit mascon amplitudes to the Stokes coefficients (e.g., Schrama and Wouters, 2011; Tiwari et al., 2009), a process that is similar in spirit to fitting mascons to level 1 data (see preceding text). In this method, the region of interest, and often adjacent regions as well, is divided into subregions (mascons), and each mascon's mass is determined by fitting all mascons simultaneously to the Stokes coefficients. The mascon solution for a subregion can sometimes be a better representation of the true spatial average for that subregion than an unscaled averaging kernel solution. Even a mascon solution, though, has contributions from mass outside the mascon and does not sample the interior mass uniformly (Jacob et al., 2012; Tiwari et al., 2009).

3.08.3 Applications

Time-variable satellite gravity measurements can be used to address a wide variety of problems, from across a broad spectrum of the Earth sciences. Any geophysical process that causes a significant redistribution of mass over scales of hundreds of kilometers or more is a possible target. In the succeeding text, we discuss some of the more prominent geophysical processes that have been studied with GRACE and illustrate the discussion by referring to a few representative publications. We make no attempt to present a complete publication review, since as of this writing (winter 2013), there have been over 2000 GRACE-related publications (for a continually updated list, see the publication page on the GRACE Tellus website: http://wwwgrace.jpl.nasa.gov). Instead, we include only a very few references, with a focus on early papers for each type of application.

3.08.3.1 Hydrology

The largest amplitude and most varied time-dependent signals are related to water storage variability on land. Figure 2(b), for example, shows that the annually varying signals on land are much larger than those in the ocean. (Before generating Figures 2, 8, and 12, the modeled oceanic contributions that were removed from the level 1 data before constructing the monthly gravity fields were added back to those fields.) When water is placed on land, a sizable fraction often stays there for some time, either infiltrating into the soil or remaining on the surface as water or snow. But when a parcel of water is placed on the ocean, its natural tendency is to flow away. Note that the largest features evident in Figure 2(b) are easily recognized: for example, heavy rainfall regions near the equator, the Southeast Asian monsoon, and the seasonal snow cycle in Eurasia and northern North America. A higher-resolution (300 km Gaussian smoothing) example is shown in Figure 8. Features clearly evident include wet regions in Central America, the heavy mountain snows that stretch from southern Alaska down to the Central Rocky Mountains, the desert region of the southwest United States and northern Mexico, the region of high-precipitation running from the lower Mississippi River basin up to Kentucky, and the high-precipitation region along the upper Saint Lawrence River.

Time-variable gravity measurements are sensitive to the total water storage integrated through the entire water column (see [7]). This includes water and snow on the surface and

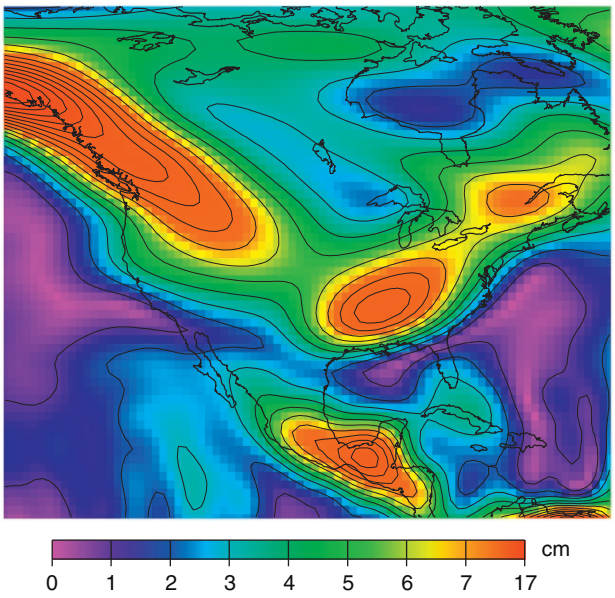


Figure 8 The amplitude of the annually varying mass signal, in units of centimeters of water thickness, recovered from GRACE during April 2002–September 2012. The results have been postprocessed to reduce noise (Swenson and Wahr, 2006a) and smoothed with a 300 km Gaussian. The solid lines represent contour intervals of 1 cm.

water in both the soil and subsoil layers. The measurements cannot distinguish between these storage components, but can recover only the sum. This hydrologic product is unique, both in its sensitivity to subsoil water storage and in its ability to recover results at large spatial scales. Other types of satellite-based instruments, either already on orbit or still in the planning stage, can detect water stored within the upper few centimeters of the soil or can monitor surface water. But time-variable gravity missions provide the only available means of monitoring deeper water storage from space. Ground-based observations from such things as soil moisture probes and the monitoring of well levels can provide information on subsurface storage at individual points. But probably, no region in the world has a dense enough observational network to provide total water storage at scales of a few hundred kilometers with the accuracy of GRACE.

3.08.3.1.1 Comparing with land surface models

Water storage estimates obtained from time-variable gravity are of potential value both as stand-alone quantities and when used in combination with other data types. As an end product, they can be compared with the total water storage predicted by land surface models, to help assess and improve those models (e.g., Andersen and Hinderer, 2005; Andersen et al., 2005; Frappart et al., 2006; Hinderer et al., 2006; Nakaegawa, 2006; Neumeyer et al., 2006; Niu and Yang, 2006; Ramillien et al., 2005; Schmidt et al., 2006b; Seo et al., 2006; Swenson and Milly, 2006; Syed et al., 2008), and with soil moisture, groundwater, and/or snow mass measurements to help validate and understand those measurements (Frappart et al., 2006; Swenson et al., 2006; Yeh et al., 2006). For example, Figure 9 (Sean Swenson, personal communication) shows comparisons between GRACE water storage estimates and those predicted by the GLDAS/Noah water storage model (Rodell et al., 2004a), for three river basins. The model includes contributions from soil moisture and snow, but does not include surface water and does not contain a groundwater layer. The GRACE error bars are defined so that if the disagreement for any month is larger than the error bars, we can be 68.3% confident that it is the model that is in error (Wahr et al., 2006).

The agreement is good for the Mississippi, which is reassuring given the high density of observations used to improve the atmospheric forcing fields in that region. The GRACE results do show somewhat larger long-period variability than GLDAS, possibly because of the lack of a groundwater store in the model. For the Amazon, the phase of the model tends to lead the phase of GRACE, and the amplitudes of the GLDAS seasonal variability are clearly too small. These discrepancies are most likely due to the absence of surface water storage in the model. These same types of discrepancies also occur, though to a notably smaller extent, for the Yenisey (in northern Siberia). In this case, the discrepancies could partly be due to the absence of surface water in GLDAS, but it could also partly be caused by difficulties in reproducing seasonal variations in snow mass. Comparisons like these can provide an indication of where model improvements are necessary. GRACE-based water storage estimates are now beginning to be assimilated directly into hydrology models (e.g., Forman et al., 2012; Li et al., 2012).

3.08.3.1.2 Floods and droughts

GRACE water storage results are proving useful for helping both to monitor droughts (e.g., Houborg et al., 2012) and to assess the risk of large-scale flooding (e.g., Reager and Famiglietti, 2009). The knowledge of the timing and severity of a drought and estimates of the probability of regional flooding can be exceedingly helpful to water managers, emergency planners, and others charged with mitigating the effects of natural disasters. GRACE is unique in its ability to provide information on large-scale changes in total water storage, including groundwater. Changes in subsurface water storage and how those changes compare with past variability are valuable diagnostic indicators of present conditions and can be helpful for forecasting purposes.

3.08.3.1.3 Anthropogenic effects and sea level contributions

One application of GRACE water storage estimates as a stand-alone product, is the general issue of hydrologic contributions to sea level change: What regions are important contributors and over what timescales? The results shown in Figure 9, for example, can be loosely interpreted as the contributions to global sea level change from those river basins (after scaling by the ratio of the land area to the area of the ocean and reversing the sign), though the connection is not that simple, of course, since the water that leaves a river basin does not necessarily go directly into the ocean.

The variability evident in Figure 9 is mostly seasonal. Of more relevance to the issue of rising sea level would be regions that display linear trends. Trends can be an indication of anthropogenic influence. Groundwater is particularly susceptible to

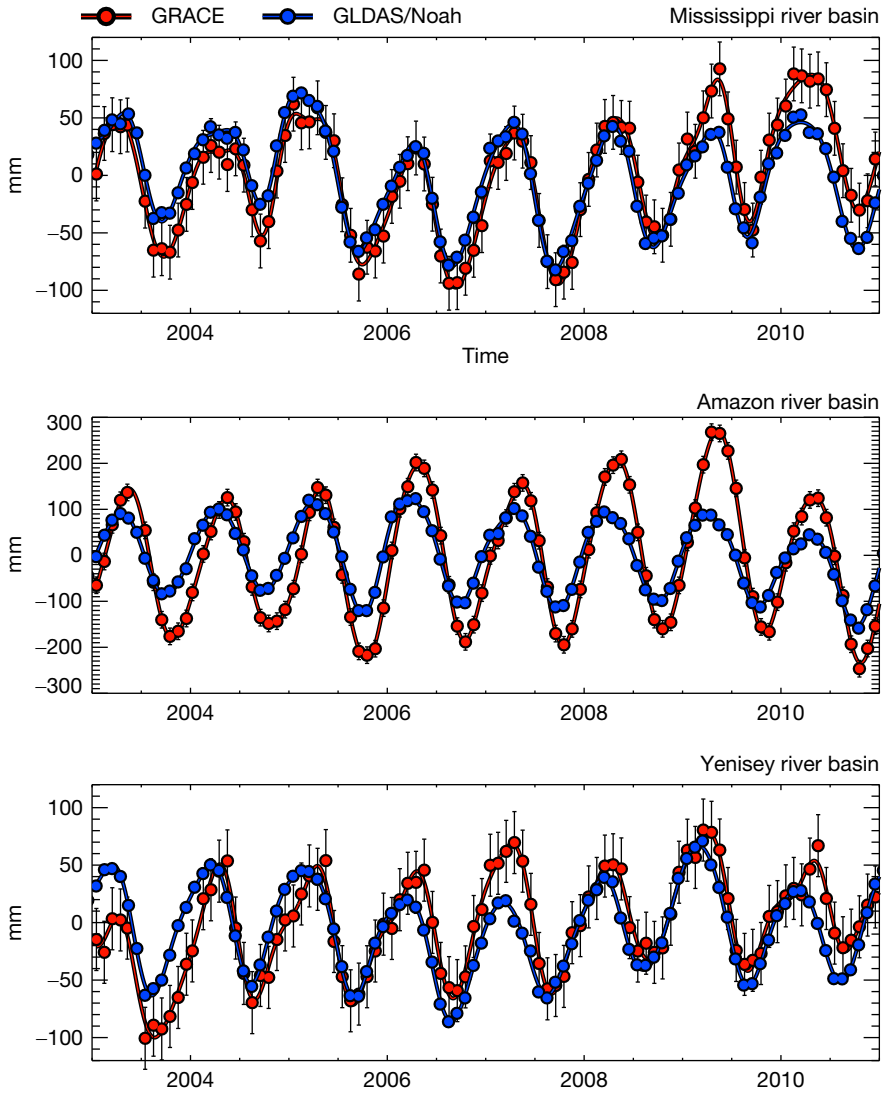


Figure 9 Water storage results for three river basins, obtained using specially constructed averaging kernels for those basins. GRACE results, with their 68.3% confidence limits, are shown in red (postprocessed to reduce noise, as described by Swenson and Wahr, 2006a). Results from the GLDAS/Noah land surface model (Rodell et al., 2004a) are in blue. Figure provided by Sean Swenson.

anthropogenic changes, both negative and positive, for example, aquifer pumping to obtain water for agricultural and urban use and groundwater infiltration from irrigation. Because few large-scale land surface models include groundwater storage and fewer still include anthropogenic effects, contributions such as these can usually not be extracted from models. Time-variable satellite gravity measurements offer a means of monitoring this variability (e.g., Famiglietti et al., 2011; Rodell et al., 2009; Scanlon et al., 2012; Tiwari et al., 2009).

3.08.3.1.4 Precipitation (P) minus evapotranspiration (ET)

P and *ET* have an important impact on climate, because their difference largely determines the exchange of mass and latent heat between the atmosphere and underlying Earth. Estimates of *P*−*ET* can be obtained from atmospheric models using moisture flux convergence parameters (e.g., Trenberth et al., 2006). Alternatively, for a land surface (hydrology) model, *P* and *ET* are typically computed using a water and energy balance approach (Roads et al., 2003). These types of models are among the best available tools for making long-range predictions of both natural and anthropogenic climate variabilities. However, because of the difficulty of obtaining relevant measurements using traditional methods, it has proven difficult to assess these model components, particularly at the synoptic scales that characterize the most energetic atmospheric disturbances. At seasonal and longer time periods, it is often assumed that storage changes are negligible and that, therefore, *P*−*ET* should balance the discharge. A model's ability to achieve this balance is sometimes used to assess the accuracy of its *P*−*ET* estimates (Gutowski et al., 1997; Roads, 2002). But water storage changes certainly do exist (e.g., see **Figures 2(b), 8**, and **9**), and at seasonal periods are typically of the same order as the discharge.

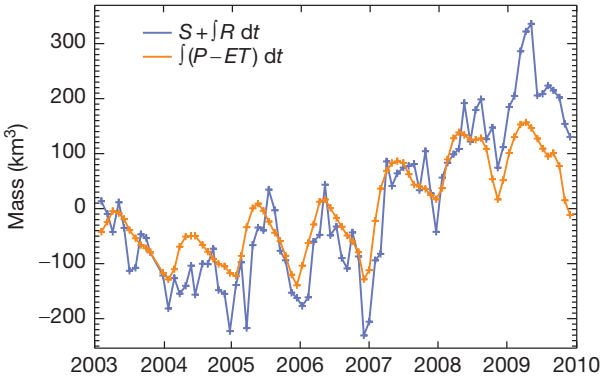


Figure 10 The blue line is the sum of GRACE water storage results, S, and the accumulated runoff, $\int dt R(t)$, for the Lena River basin in Eurasia and so is an estimate of the accumulated precipitation minus evapotranspiration, $\int dt[P-ET](t)$, within the Lena River basin. The orange line is $\int dt[P-ET](t)$ as inferred from moisture convergence results based on ERA-Interim atmospheric forecast fields. Figure after Velicogna et al. (2012); data provided by Isabella Velicogna.

Time-variable gravity offers an additional opportunity for determining $P-ET$ (see Rodell et al., 2004b; Swenson and Wahr, 2006b; Velicogna et al., 2012). The water budget equation is

$$S(t)-S(t_0)=\int_{t_0}^{t}[P(t)-ET(t)-R(t)]\,dt \qquad [25]$$

where S is total water storage and R is discharge. Time-variable gravity measurements can be used to estimate S in a river basin. If the river that drains that basin is gauged, then the discharge can be measured and so $P-ET$ can be determined. As an example, **Figure 10** (after Figure 3 of Velicogna et al., 2012; data provided by Isabella Velicogna) compares $\int_{t_0}^{t}[P-ET]dt$ estimates for the Lena River in Siberia constructed using GRACE and river discharge measurements (i.e., $S(t)+\int_{t_0}^{t}R\,dt$) with atmospheric model estimates from the ERA-Interim reanalysis atmospheric fields (Simmons et al., 2007). Clearly, this atmospheric model does a good job at reproducing $P-ET$ in this basin.

As a variation of this application, suppose atmospheric models are believed to accurately predict $P-ET$ within some river basin. Time-variable gravity estimates of S can then be used in [25], along with the $P-ET$ results, to estimate the river discharge (Landerer et al., 2010; Syed et al., 2005). This offers a means of determining discharge for rivers that are not adequately gauged.

3.08.3.2 Cryosphere

One of the most important consequences of rising global temperatures is increased global sea levels caused by accelerated mass loss from the Antarctic and Greenland ice sheets. There is enough frozen water in those ice sheets to raise the world's oceans by 70 m if they melted completely. Thus, even a relatively small change in ice sheet mass can have a significant impact on sea level. Accurate monitoring of ice sheet variability provides information about the present state of mass imbalance of the ice sheets, which is essential for improving predictions of future changes. Several satellite geodetic techniques have been developed over the past few decades that have led to significant improvements in ice sheet monitoring. Among these are radar- and laser-altimeter measurements of ice sheet elevations, radar measurements of velocities and thinning rates of outlet glaciers and ice streams (combined with models of surface accumulation, sublimation, and melting), and GRACE gravity measurements. Each of these techniques has its own characteristic strengths and weaknesses.

Satellite gravity has two distinct advantages over other techniques. First, gravity measurements provide a direct estimate of mass, which is obviously the most relevant quantity for understanding mass imbalance. Other methods do not determine mass loss directly, but rely on independent assumptions to relate measured quantities to mass. Second, gravity signals at the altitude of a satellite are caused by mass variations averaged over a broad region of the underlying surface, not just at the point directly beneath the satellite. Thus, satellite gravity inherently averages over large regions. Other methods tend either to sample an ice sheet at relatively small nonoverlapping footprints or to miss entire regions, so that their estimates of total mass imbalance are subject to interpolation and extrapolation errors.

Time-variable gravity has weaknesses, of course. For one thing, it cannot provide small-scale resolution and so has trouble isolating the exact location of a mass anomaly. For another, time-variable gravity estimates are particularly sensitive to PGR errors. Both Antarctica and Greenland experienced significant melting at the end of the last ice age, and the underlying Earth is still rebounding. This rebound affects altimeter estimates of ice sheet thickness change: if the crust rises (or falls), the ice sheet's surface will rise (fall) along with it, and so the altimeter data will imply that the ice sheet is getting thicker (thinner). It affects satellite gravity estimates because it produces a gravity signal that is inseparable from the gravity signal caused by the ongoing ice change. Because rock deep in the Earth's mantle is four to five times as dense as ice, the impact of PGR on gravity is relatively four to five times as large as its impact on altimeter estimates. If the Earth's surface uplifts by 1 cm, the altimeter sees the ice sheet surface rise by 1 cm. But a gravity measurement sees a gravity signal that is the equivalent of the signal from 4 to 5 cm of ice. Thus, although PGR models are usually used to remove the PGR signals from both altimeter and gravity measurements, any residual errors in those models cause more problems for gravity than for altimetry. Ultimately, the best approach may be to combine time-variable gravity and altimeter estimates, as well as GPS observations of vertical crustal motion where available, to reduce the PGR errors in both techniques (Riva et al., 2009; Velicogna and Wahr, 2002).

There have been numerous GRACE estimates of Antarctic and Greenland mass loss, starting with Velicogna and Wahr (2006a,b), Chen et al. (2006b,c), and Luthcke et al. (2006). A recent multiauthor study that develops and compares GRACE ice sheet results using a number of analysis methods

(and that also compares GRACE results with those based on other satellite techniques) is Shepherd et al. (2012).

GRACE has also been used to estimate ice loss from ice caps and mountain glacier systems (see, e.g., Chen et al., 2006d, 2007a; Gardner et al., 2011; Jacob et al. 2012; Luthcke et al., 2008; Tamisiea et al., 2005). Those regions are smaller than the Greenland and Antarctic ice sheets and so are harder to resolve. In addition, since most of the important mountain glacier systems are adjacent to non-ice-covered land, GRACE results for those regions can be sensitive to changes in liquid water stored on that land. In such cases, hydrology model output must be used to remove the contamination from the water storage signal. Despite these difficulties, GRACE has the same significant advantages for monitoring these ice-covered regions, as it has for the polar ice sheets: a direct estimate of mass change and complete spatial coverage.

3.08.3.3 Solid Earth

Although the Earth's mean gravity field is caused almost entirely by mass within the solid Earth, changes in the distribution of that mass generally occur too slowly or produce gravity signals that are too small or too localized, to be practical targets of time-variable satellite gravity studies. The most notable exceptions are PGR and very large earthquakes.

The PGR signal over Canada is clearly visible in the GRACE data and has proven useful in helping to constrain the Earth's viscosity profile (e.g., Paulson et al., 2007; Tamisiea and Davis, 2006). **Figure 11(a)**, for example, shows the best-fitting linear trend in surface mass, smoothed with a 400 km Gaussian. **Figure 11(b)** shows the expected PGR signal over that same region, computed by A et al. (2013) using the ICE-5G ice deglaciation model and VM2 viscosity profile (Peltier, 2004). There is good general agreement with the GRACE observations over Hudson Bay, as evidenced in **Figure 11(c)**, which shows that after removing the ICE-5G results from GRACE, the Hudson Bay anomaly is dramatically reduced. The remaining negative anomaly over southern Alaska has been interpreted as the effects of shrinking glaciers (Chen et al., 2006d; Luthcke et al., 2008; Tamisiea et al., 2005). PGR signals in Antarctica and Greenland are harder to recover using GRACE, due to the problems of separating those signals from present-day mass variability of the ice sheets, as mentioned earlier in the text.

Few earthquakes have been unambiguously detected using GRACE data. Exceptions are the 2004 Sumatra–Andaman, 2010 Maule (Chile), and 2011 Tohoku–Oki (Japan) events (see, e.g., Cambiotti and Sabadini, 2012; Chen et al., 2007b; Han et al., 2006, 2008, 2010; Matsuo and Heki, 2011; Panet et al., 2007; Wang et al., 2012), all of which were unusually energetic earthquakes. Both co- and postseismic signals have been studied with GRACE. A coseismic temporal signature appears as a step function in the GRACE record. Postseismic deformation, which appears as a slowly decaying signal after an earthquake, is arguably of even more interest since postseismic displacements cannot be detected using conventional seismometers. GPS measurements can also be used to detect co- and postseismic deformation, though of course they detect surface displacements rather than the gravity signal. Dense GPS networks have better spatial resolution than GRACE. But GRACE has the advantage of being able to detect signals over the ocean, where there are unlikely to be GPS receivers. And many of the world's most active plate boundaries are located beneath or adjacent to the ocean.

There is, in addition, an indirect way in which time-variable satellite gravity measurements can contribute to solid Earth studies. Global Positioning System (GPS) observations are widely used to monitor tectonic displacements of the Earth's surface. But the Earth's surface can deform in response to surface loading, as well. Load deformation is a source of noise for tectonic applications. It can be especially troublesome for campaign-style GPS observing programs, in which a site might be occupied for a few days and then not reoccupied for perhaps several years. In that situation, seasonal and other short-period loading can alias into apparent long-term variability.

Time-variable satellite gravity observations can be used to model and remove the large-scale component of the load deformation and so to reduce this source of noise. The surface mass variability recovered from those observations can be convolved with solid Earth Green's functions to estimate the loading (e.g., Davis et al., 2004; Fu et al., 2012; King et al., 2006; Tregoning et al., 2009; van Dam et al., 2007).

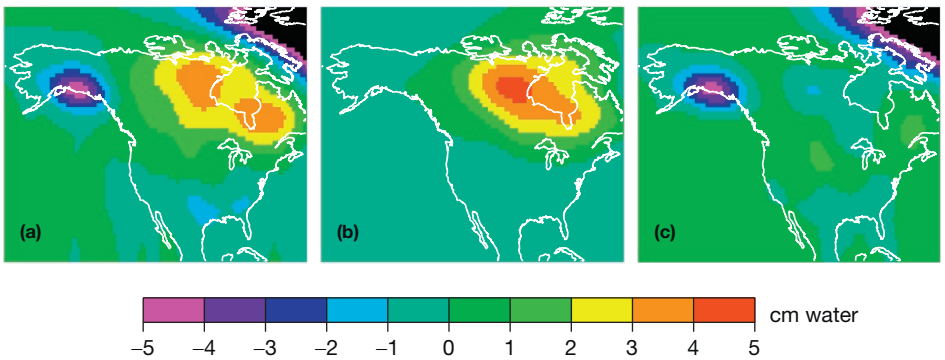


Figure 11 (a) The best-fitting linear trend for the GRACE fields between April 2002 and September 2012, smoothed with a 400 km Gaussian smoothing function. Units are centimeter per year of water thickness. (b) PGR predictions from A et al. (2013), based on the ICE-5G/VM2 ice deglaciation and viscosity models (Peltier, 2004). (c) The difference between (a) and (b).

3.08.3.4 Oceanography

The time-variable mass signal in the ocean is small compared to that from land, as can be seen for the annual cycle from Figure 2(b). Bottom pressure variability is not dominated by an annual signal to the same extent as land water storage. But even when all temporal variations are included, ocean mass fluctuations are still relatively small. Figure 12, for example, shows that the rms surface mass variability over the oceans, smoothed with a 750 km Gaussian, is typically only 2–3 cm or less. Presumably, the large red regions along coastlines mostly reflect the effects of the much larger land water signals leaking into the ocean estimates. This illustrates the danger of using time-variable gravity to study the ocean near the coast. This leakage can be reduced by decreasing the smoothing radius or by removing the water storage output of a land surface model or by using an iterative procedure to first solve for the contributions from the hydrology signal (Wahr et al., 1998).

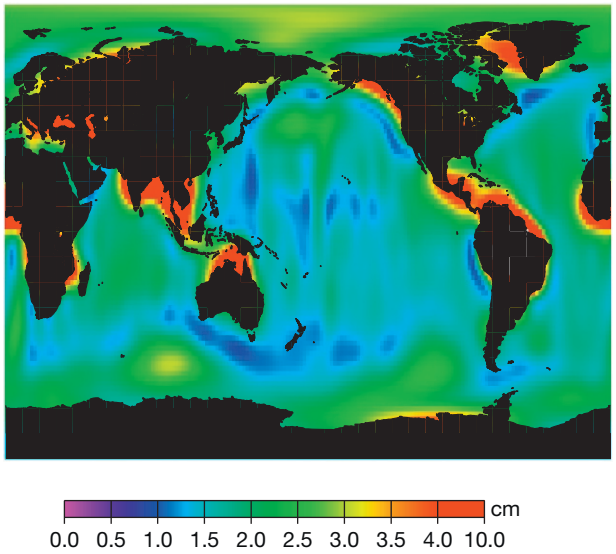


Figure 12 The rms about the mean of ocean surface mass variability deduced from GRACE for April 2002–September 2012, smoothed with a 750 km Gaussian.

Still, ocean mass signals have been clearly identified in GRACE data. One of the direct oceanographic applications of these mass estimates is to combine them with sea surface height measurements from altimetry, to separately estimate steric and nonsteric contributions to sea surface variability. A satellite radar altimeter monitors sea surface height along its ground track. Suppose the altimeter detects a sea surface rise in some region. The altimeter data cannot determine whether the rise was due to increased water mass in the region or to the water becoming warmer (and/or less salty) and expanding. Change in volume (i.e., 'steric' changes) does not cause a change in gravity, and so satellite gravity measurements detect only the nonsteric contributions. The steric contributions are then the difference between the altimeter and time-variable gravity results.

Figure 13 (results updated from Willis et al. (2008) and provided by Don Chambers) shows how well this technique works on a global scale (see also Chambers 2006; Lombard et al., 2007). The red curve is the total ocean mass deduced from GRACE. The blue curve shows an altimetric estimate of sea surface height, corrected for steric effects using temperature and salinity measurements from the Argo array of profiling floats. The agreement is excellent, not only for the seasonal terms but also for the longer-term variability.

If the time-variable steric signal can be estimated for some region by differencing altimeter and GRACE mass results, it is possible to recover changes in the heat content of that region, as

$$\Delta H = \frac{\rho c_p}{\alpha}\left(\Delta\eta - \frac{1}{\rho_0}\Delta\sigma\right) \qquad [26]$$

where ΔH is the change in ocean heat storage, ρ is the ocean density, c_p and α are the heat capacity and thermal expansion coefficient of sea water, respectively, $\Delta\sigma$ is the change in mass estimated from time-variable gravity, and $\Delta\eta$ is the change in sea

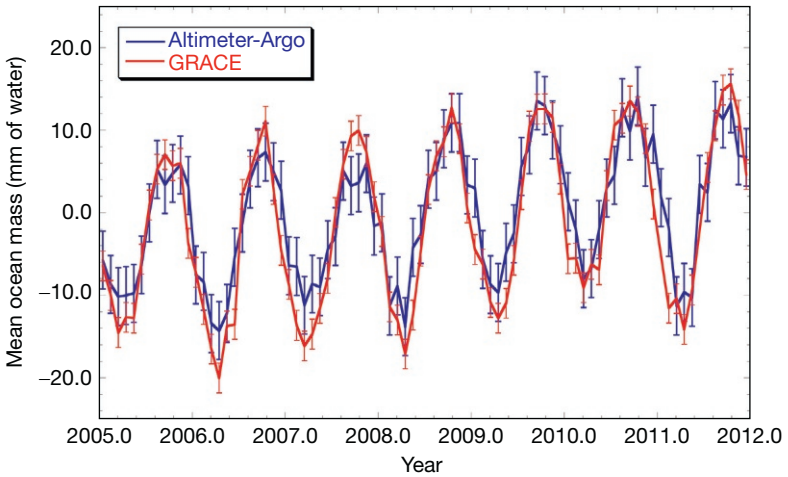


Figure 13 Global ocean mass variability, expressed as mean ocean thickness, measured by GRACE and estimated from steric-corrected altimetry (updated from Willis et al., 2008). GRACE results are computed from the CSR Release-05 gravity fields, the altimetry is from Jason-1/Jason-2, and the steric signal is computed from mapped Argo temperature/salinity data. The average is calculated between 60°S and 60°N, because of the limitation of the altimetry/Argo coverage. The uncertainty bars are one standard error. Plot provided by Don Chambers.

surface height measured by the altimeter (Jayne et al., 2003). This result, which extends the methodology of Chambers et al. (1997) to include mass variability, assumes that α is independent of depth and that the effects of salinity variations either are negligible or can be independently modeled and removed.

The exchange of heat between the ocean and atmosphere is one of the most significant examples of energy transfer within the Earth's climate system. Because of the large heat capacity of water, the ocean can store enormous amounts of energy. Therefore, it can act not only as a moderator of climate extremes but also as an energy source for severe storms. Knowledge of the ocean's time-varying heat storage is of considerable importance for such things as climate change prediction, long-range weather forecasting, and hurricane strength prediction. Because of its global coverage, the combination of GRACE with altimetry provides a means of assessing and extending float-based maps of the ocean's time-varying heat storage.

Time-variable mass estimates can be used for other types of oceanographic applications, as well. Surface mass anomalies, $\Delta\sigma$, are proportional to variations in ocean bottom pressure

$$\Delta P_{\text{bott}}(\theta, \phi) = g\Delta\sigma(\theta, \phi) \qquad [27]$$

where $\Delta\sigma$ [7] is integrated from the bottom of the ocean to the top of the atmosphere. Thus, time-variable gravity over the ocean provides estimates of seafloor pressure variability at the spatial and temporal resolution of the gravity measurements. GRACE, for example, can provide monthly seafloor pressure maps at scales of several hundred kilometers and greater. These can be used to assess and improve oceanographic models (Bingham and Hughes, 2006; Condi and Wunsch, 2004; Zlotnicki et al., 2007) and to compare with measurements from bottom pressure recorders (Kanzow et al., 2005; Morison et al., 2007) to separate the effects of regional and local signals.

The bottom pressure estimates can also be combined with the geostrophic assumption (which assumes a balance between pressure and Coriolis forces) to determine changes in deep ocean velocities at the temporal and spatial resolutions of the gravity field observation. For GRACE, this means the results are averaged over scales of several hundred kilometers or more. This large a spatial scale can make it difficult to apply the geostrophic assumption at the seafloor in the presence of short-scale topography. However, simulations (Wahr et al., 2002) have shown that pressure variability at the seafloor is about the same as at 2 km depth or at even shallower depths in many cases. Thus, the inferred currents can be interpreted to a high degree of accuracy in terms of the variability of currents at 2 km depth.

3.08.4 The Future

The launch of GRACE in the spring of 2002 signaled the start of a new era in satellite geodesy. Although satellite laser ranging has long been providing time-variable gravity results for a small handful of global-scale spherical harmonic components, it is the much higher spatial resolution provided by GRACE that allows the wide variety of applications described in this chapter. Changes in continental water storage, in ice sheet mass, in ocean bottom pressure, and in the Earth's internal density distribution – any process, anywhere on the globe, that involves significant redistribution of mass at scales of a few hundred kilometers and greater – are potential targets for GRACE. And the analysis methods used to recover mass variability from GRACE fields are straightforward and easy to implement.

GRACE, of course, will eventually fail. It was designed for only a 5-year lifetime, but as of this writing (winter 2013), it has lasted for over a decade, although with some degradation especially to the battery performance. A follow-on mission, funded by NASA and the German BMBF, is in development, with a projected 2017 launch (Watkins and Flechtner, 2012). This mission will be similar to GRACE, with several spacecraft and instrumentation upgrades. Although it will continue to use a microwave interferometer as its primary method of monitoring the satellite-to-satellite distance, it will also include a prototype laser interferometer as a technical demonstration payload. The mission will permit an extension of the types of measurements GRACE has been providing since 2002, with some improvement in accuracy and resolution.

References

A G, Wahr J, and Zhong S (2013) Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: An application to Glacial Isostatic Adjustment in Antarctica and Canada. *Geophysical Journal International* 192: 557–572. http://dx.doi.org/10.1093/gji/ggs030.

Andersen OB and Hinderer J (2005) Global inter-annual gravity changes from GRACE: Early results. *Geophysical Research Letters* 32(1): L01402.

Andersen OB, Seneviratne SI, Hinderer J, and Viterbo P (2005) GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophysical Research Letters* 32(18): L18405.

Benjamin D, Wahr J, Ray RD, Egbert GD, and Desai SD (2006) Constraints on mantle anelasticity from geodetic observations, and implications for the J_2 anomaly. *Geophysical Journal International* 165: 3–16.

Bingham RJ and Hughes CW (2006) Observing seasonal bottom pressure variability in the North Pacific with GRACE. *Geophysical Research Letters* 33(8): L08607.

Cambiotti G and Sabadini R (2012) A source model for the great 2011 Tohoku earthquake ($M_w = 9.1$) from inversion of GRACE gravity data. *Earth and Planetary Science Letters* 335: 72–79. http://dx.doi.org/10.1016/j.epsl.2012.05.002.

Chambers DP (2006) Observing seasonal steric sea level variations with GRACE and satellite altimetry. *Journal of Geophysical Research: Oceans* 111: 3010.

Chambers DP, Tapley BD, and Stewart RH (1997) Long-period ocean heat storage rates and basin-scale heat fluxes from TOPEX. *Journal of Geophysical Research: Oceans* 102: 10525–10533.

Chao BF (1994) The geoid and Earth rotation. In: Vanicek P and Christou N (eds.) *Geoid and Its Geophysical Interpretations*, pp. 285–298. Boca Raton, FL: CRC Press.

Chao BF and Au AY (1991) Temporal variation of the Earth's low-degree zonal gravitational field caused by atmospheric mass distribution, 1980–1988. *Journal of Geophysical Research* 96(B4): 6569–6575.

Chao BF and Gross RS (1987) Changes in the Earth's rotation and low-degree gravitational field induced by earthquakes. *Geophysical Journal of the Royal Astronomical Society* 91: 569–596.

Chen JL, Tapley BD, and Wilson CR (2006) Alaskan mountain glacial melting observed by satellite gravimetry. *Earth and Planetary Science Letters* 248(1–2): 368–378.

Chen JL, Wilson CR, Blankenship DD, and Tapley BD (2006) Antarctic mass rates from GRACE. *Geophysical Research Letters* 33(11): L11502.

Chen JL, Wilson CR, and Seo KW (2006) Optimized smoothing of gravity recovery and climate experiment (GRACE) time-variable gravity observations. *Journal of Geophysical Research* 111: B06408.

Chen JL, Wilson CR, and Tapley BD (2006) Satellite gravity measurements confirm accelerated melting of Greenland ice sheet. *Science* 313: 1958–1960.

Chen JL, Wilson CR, Tapley BD, Blankenship DD, and Ivins ER (2007) Patagonia icefield melting observed by gravity recovery and climate experiment (GRACE). *Geophysical Research Letters* 34: L22501.

Chen JL, Wilson CR, Tapley BD, and Grand S (2007) GRACE detects coseismic and postseismic deformation from the Sumatra-Andaman earthquake. *Geophysical Research Letters* 34: L13302. http://dx.doi.org/10.1029/2007GL030356.

Cheng MK, Shum CK, and Tapley BD (1997) Determination of long-term changes in the Earth's gravity field from satellite laser ranging observations. *Journal of Geophysical Research* 102(B10): 22377–22390.

Cheng MK and Tapley BD (1999) Seasonal variations in low degree zonal harmonics of the Earth's gravity field from satellite laser ranging observations. *Journal of Geophysical Research* 104(B2): 2667–2681.

Cheng MK, Tapley BD, and Ries JC (2013) Deceleration in the Earth's oblateness. *Journal of Geophysical Research*. http://dx.doi.org/10.1002/jgrb.50058.

Condi F and Wunsch C (2004) Measuring gravity field variability, the geoid, ocean bottom pressure fluctuations, and their dynamical implications. *Journal of Geophysical Research* 109(C2): C02013.

Cox CM and Chao BF (2002) Detection of a large-scale mass redistribution in the terrestrial system since 1998. *Science* 297: 831–832.

Davis JL, Elosegui P, Mitrovica JX, and Tamisiea ME (2004) Climate-driven deformation of the solid Earth from GRACE and GPS. *Geophysical Research Letters* 31(24): L24605.

Dickey JO, Bentley CR, Bilham R, et al. (1997) *Satellite gravity and the geosphere*. *National Research Council Report*, p. 112. National Academy Press.

Dickey JO, Marcus SL, de Viron O, and Fukumori I (2002) Recent Earth oblateness variations: Unraveling climate and postglacial rebound effects. *Science* 298: 1975–1977.

Dong D, Gross RS, and Dickey JO (1996) Seasonal variations of the Earth's gravitational field: An analysis of atmospheric and oceanic tidal excitation. *Geophysical Research Letters* 23: 725–728.

Dziewonski A and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

Famiglietti JS, Lo M, Ho SL, et al. (2011) Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters* 38. http://dx.doi.org/10.1029/2010GL046442.

Farrell WE (1972) Deformation of the Earth by surface loading. *Reviews of Geophysics* 10: 761–797.

Forman BA, Reichle RH, and Rodell M (2012) Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resources Research* 48. http://dx.doi.org/10.1029/2011WR011239.

Frappart F, Ramillien G, Biancamaria S, Mognard NM, and Cazenave A (2006) Evolution of high-latitude snow mass derived from the GRACE gravimetry mission (2002–2004). *Geophysical Research Letters* 33(2): L02501.

Fu YN, Freymueller JT, and Jensen T (2012) Seasonal hydrological loading in southern Alaska observed by GPS and GRACE. *Geophysical Research Letters* 39: L15310. http://dx.doi.org/10.1029/2012GL052453.

Gardner AS, Moholdt G, Wouters B, et al. (2011) Sharply increased mass loss from glaciers and ice caps in the Canadian Arctic Archipelago. *Nature* 473: 357–360.

Gutowski WJ, Chen Y, and Otles Z (1997) Atmospheric water vapor transport in NCEP-NCAR reanalyses: Comparison with river discharge in the central United States. *Bulletin of the American Meteorological Society* 78: 1957–1969.

Han SC, Jekeli C, and Shum CK (2004) Time-variable aliasing effects of ocean tides, atmosphere, and continental water mass on monthly mean GRACE gravity field. *Journal of Geophysical Research* 109(B4): B04403.

Han SC, Sauber J, and Luthcke S (2010) Regional gravity decrease after the 2010 Maule (Chile) earthquake indicates large-scale mass redistribution. *Geophysical Research Letters* 37. http://dx.doi.org/10.1029/2010GL045449.

Han SC, Sauber J, Luthcke SB, Ji C, and Pollitz FF (2008) Implications of postseismic gravity change following the great 2004 Sumatra-Andaman earthquake from the regional harmonic analysis of GRACE intersatellite tracking data. *Journal of Geophysical Research* 113: B11413. http://dx.doi.org/10.1029/2008JB005705.

Han SC, Shum CK, Bevis M, Ji C, and Kuo CY (2006) Crustal dilatation observed by GRACE after the 2004 Sumatra-Andaman earthquake. *Science* 313: 658–662.

Han SC, Shum CK, Jekeli C, and Alsdorf D (2005) Improved estimation of terrestrial water storage changes from GRACE. *Geophysical Research Letters* 32(7): L11502, L07302.

Han SC, Shum CK, Jekeli C, Kuo CY, Wilson C, and Seo KW (2005) Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement. *Geophysical Journal International* 163: 18–25.

Han SC, Shum CK, and Matsumoto K (2005) GRACE observations of M-2 and S-2 ocean tides underneath the Filchner-Ronne and Larsen ice shelves, Antarctica. *Geophysical Research Letters* 32(20): L20311.

Han SC and Simons FJ (2008) Spatiospectral localization of global geopotential fields from the Gravity Recovery and Climate Experiment (GRACE) reveals the coseismic gravity change owing to the 2004 Sumatra-Andaman earthquake. *Journal of Geophysical Research: Solid Earth* 113: B01405. http://dx.doi.org/10.1029/2007JB004927.

Han D and Wahr J (1995) The viscoelastic relaxation of a realistically stratified Earth, and a further analysis of postglacial rebound. *Geophysical Journal International* 120: 287–311.

Hinderer J, Andersen O, Lemoine F, Crossley D, and Boy JP (2006) Seasonal changes in the European gravity field from GRACE: A comparison with superconducting gravimeters and hydrology model predictions. *Journal of Geodynamics* 41(1–3): 59–68.

Houborg R, Rodell M, Li BL, Reichle R, and Zaitchik BF (2012) Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research* 48: W07525. http://dx.doi.org/10.1029/2011WR011291.

Jacob T, Wahr J, Pfeffer T, and Swenson S (2012) Recent contributions of glaciers and ice caps to sea level rise. *Nature*. http://dx.doi.org/10.1038/nature10847.

Jayne SR, Wahr JM, and Bryan FO (2003) Observing ocean heat content using satellite gravity and altimetry. *Journal of Geophysical Research: Oceans* 108: 3031.

Jekeli C (1981) Alternative methods to smooth the Earth's gravity field. *Report 327*. Columbus: Department of Geodetic Science and Surveying, Ohio State University.

Jekeli C (1999) The determination of gravitational potential differences from satellite-to-satellite tracking. *Celestial Mechanics and Dynamical Astronomy* 75: 85–100.

Kanzow T, Flechtner F, Chave A, Schmidt R, and Schwintzer P (2005) Seasonal variation of ocean bottom pressure derived from Gravity Recovery and Climate Experiment (GRACE): Local validation and global patterns. *Journal of Geophysical Research* 110(C9): C09001.

King M, Moore P, Clarke P, and Lavallee D (2006) Choice of optimal averaging radii for temporal GRACE gravity solutions, a comparison with GPS and satellite altimetry. *Geophysical Journal International* 166(1): 1–11.

Knudsen P and Andersen O (2002) Correcting GRACE gravity fields for ocean tide effects. *Geophysical Research Letters* 29(8): 1178.

Landerer FW, Dickey JO, and Guentner A (2010) Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009. *Journal of Geophysical Research* 115: D23115. http://dx.doi.org/10.1029/2010JD014584.

Landerer FW and Swenson SC (2012) Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research* 48: W045531. http://dx.doi.org/10.1029/2011WR011453.

Lavallee DA, Moore P, Clarke PJ, Petrie EJ, van Dam T, and King MA (2010) J2: An evaluation of new estimates from GPS, GRACE, and load models compared to SLR. *Geophysical Research Letters* 37: L22403. http://dx.doi.org/10.1029/2010GL045229.

Lemoine FG, Kenyon SC, Factor JK, et al. (1998) *The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96*. NASA/TP-1998-20681, Goddard Space Flight Center.

Li N, Rodell M, Zaitchik BF, Reichle RH, Koster RD, and van Dam TM (2012) Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe. *Journal of Hydrology* 446: 103–115. http://dx.doi.org/10.1016/j.jhydrol.2012.04.035.

Lombard A, Garcia D, Ramillien G, et al. (2007) Estimation of steric sea level variations from combined GRACE and Jason-1 data. *Earth and Planetary Science Letters* 254: 194–202. http://dx.doi.org/10.1016/j.epsl.2006.11.035.

Luthcke SB, Arendt AA, Rowlands DD, McCarthy JJ, and Larsen CF (2008) Recent glacier mass changes in the Gulf of Alaska region from GRACE mascon solutions. *Journal of Glaciology* 54: 767–777.

Luthcke SB, Zwally HJ, Abdalati W, et al. (2006) Recent Greenland ice mass loss by drainage system from satellite gravity observations. *Science* http://dx.doi.org/10.1126/science.1130776.

Matsuo K and Heki K (2011) Coseismic gravity changes of the 2011 Tohoku-Oki earthquake from satellite gravimetry. *Geophysical Research Letters* 38. http://dx.doi.org/10.1029/2011GL049018.

Moore P, Zhang Q, and Alothman A (2005) Annual and semiannual variations of the Earth's gravitational field from satellite laser ranging and CHAMP. *Journal of Geophysical Research* 110(B6): B06401.

Morison J, Wahr J, Kwok R, and Peralta-Ferriz C (2007) Recent trends in Arctic Ocean mass distribution revealed by GRACE. *Geophysical Research Letters* 34: L07602. http://dx.doi.org/10.1029/2006GL029016.

Nakaegawa T (2006) Detectability assessment of interannual variations in terrestrial water storage from satellite gravimetry using an offline land surface model simulation. *Hydrological Processes* 20(6): 1347–1364.

Nerem RS, Eanes RJ, Thompson PF, and Chen JL (2000) Observations of annual variations of the Earth's gravitational field using satellite laser ranging and geophysical models. *Geophysical Research Letters* 27(12): 1783–1786.

Nerem RS and Wahr J (2011) Recent changes in the Earth's oblateness driven by Greenland and Antarctic ice mass loss. *Geophysical Research Letters* 38: L13501. http://dx.doi.org/10.1029/2011GL047879.

Neumeyer J, Barthelmes F, Dierks O, et al. (2006) Combination of temporal gravity variations resulting from superconducting gravimeter (SG) recordings, GRACE satellite observations and global hydrology models. *Journal of Geodesy* 79(10–11): 573–585.

Niu GY and Yang ZL (2006) Assessing a land surface model's improvements with GRACE estimates. *Geophysical Research Letters* 33(7): L07401.

Panet I, Mikhailov V, Diament M, et al. (2007) Coseismic and post-seismic signatures of the Sumatra 2004 December and 2005 March earthquakes in GRACE satellite gravity. *Geophysical Journal International* 171: 177–190. http://dx.doi.org/10.1111/j.1365-246X.2007.03525.x.

Paulson A, Zhong S, and Wahr J (2007) Inference of mantle viscosity from GRACE and relative sea level data. *Geophysical Journal International* 171: 497–508.

Peltier WR (2004) Global glacial isostasy and the surface of the ice-age Earth: The ICE-5G(VM2) model and GRACE. *Annual Review of Earth and Planetary Sciences* 32: 111–149.

Ramillien G, Frappart F, Cazenave A, and Guntner A (2005) Time variations of land water storage from an inversion of 2 years of GRACE geoids. *Earth and Planetary Science Letters* 235(1–2): 283–301.

Ray R and Luthcke S (2006) Tide model errors and GRACE gravimetry: Towards a more realistic assessment. *Geophysical Journal International* 167: 1055–1059. http://dx.doi.org/10.1111/j.1365-246X.2006.03229.x.

Reager JT and Famiglietti JS (2009) Global terrestrial water storage capacity and flood potential using GRACE. *Geophysical Research Letters* 36: L23402. http://dx.doi.org/10.1029/2009GL040826.

Reigber C, Luhr H, and Schwintzer P (2002) CHAMP mission status. *Advances in Space Research* 30(2): 129–134.

Riva REM, Gunter BC, Urban TJ, et al. (2009) Glacial isostatic adjustment over Antarctica from combined ICESat and GRACE satellite data. *Earth and Planetary Science Letters* 288(3–4): 516–523.

Roads J (2002) Closing the water budget. *GEWEX News* 12(1): 1–6.

Roads J, Lawford R, Bainto E, et al. (2003) GCIP water and energy budget synthesis (WEBS). *Journal of Geophysical Research* 108(16): 8609. http://dx.doi.org/10.1029/2002JD002583.

Rodell M, Velicogna I, and Famiglietti JS (2009) Satellite-based estimates of groundwater depletion in India. *Nature* 460: 999–1002. http://dx.doi.org/10.1038/nature08238.

Rodell M, Houser PR, Jambor U, et al. (2004a) The global land data assimilation system. *Bulletin of the American Meteorological Society* 85: 381–394.

Rodell M, Famiglietti JS, Chen J, et al. (2004b) Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophysical Research Letters* 31: L20504. http://dx.doi.org/10.1029/2004GL020873.

Rowlands DD, Luthcke SB, Klosko SM, et al. (2005) Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. *Geophysical Research Letters* 32(4): L04310.

Rubincam DP (1984) Postglacial rebound observed by Lageos and the effective viscosity of the lower mantle. *Journal of Geophysical Research* 89: 1077–1088.

Scanlon BR, Longuevergne L, and Long D (2012) Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. *Water Resources Research* 48. http://dx.doi.org/10.1029/2011WR011312.

Schmidt M, Han SC, Kusche J, Sanchez L, and Shum CK (2006) Regional high-resolution spatiotemporal gravity modeling from GRACE data using spherical wavelets. *Geophysical Research Letters* 33(8): L08403.

Schmidt R, Schwintzer P, Flechtner F, et al. (2006) GRACE observations of changes in continental water storage. *Global and Planetary Change* 50(1–2): 112–126.

Schrama EJO (2004) Impact of limitations in geophysical background models on follow-on gravity missions. *Earth, Moon, and Planets* 94: 143–163.

Schrama EJO and Wouters B (2011) Revisiting Greenland ice sheet mass loss observed by GRACE. *Journal of Geophysical Research: Solid Earth* 116: B02407. http://dx.doi.org/10.1029/2009JB006847.

Seo KW and Wilson CR (2005) Simulated estimation of hydrological loads from GRACE. *Journal of Geodesy* 78: 442–456. http://dx.doi.org/10.1007/s00190-004-0410-5.

Seo KW, Wilson CR, Famiglietti JS, Chen JL, and Rodell M (2006) Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research* 42(5): W05417.

Shepherd A, Ivins ER, Geruo A, et al. (2012) A reconciled estimate of ice-sheet mass balance. *Science* 338(6111): 1183–1189. http://dx.doi.org/10.1126/science.1228102.

Simmons AJ, Uppala S, Dee D, and Kobayashi S (2007) ERA-interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter* 110: 1–52.

Song YT and Zlotnicki V (2004) Ocean bottom pressure waves predicted in the tropical Pacific. *Geophysical Research Letters* 31(5): L05306.

Swenson S, Chambers D, and Wahr J (2008) Estimating geocenter variations from a combination of GRACE and ocean model output. *Journal of Geophysical Research* 113: B08410. http://dx.doi.org/10.1029/2007JB005338.

Swenson S and Milly PCM (2006) Climate model biases in seasonality of continental water storage recovered by satellite gravimetry. *Water Resources Research* 42: W03201. http://dx.doi.org/10.1029/2005WR004628.

Swenson S and Wahr J (2002a) Estimated effects of the vertical structure of atmospheric mass on the time-variable geoid. *Journal of Geophysical Research* 107(B9): 2194. http://dx.doi.org/10.1029/2001JB000515.

Swenson S and Wahr J (2002b) Methods for inferring regional surface-mass anomalies from GRACE measurements of time-variable gravity. *Journal of Geophysical Research* 107(B9): 2193.

Swenson S and Wahr J (2006a) Post-processing removal of correlated errors in GRACE data. *Geophysical Research Letters* 33: L08402. http://dx.doi.org/10.1029/2005GL025285.

Swenson S and Wahr J (2006b) Estimating large-scale precipitation minus evapotranspiration from GRACE satellite gravity measurements. *Journal of Hydrometeorology* 7: 252–270.

Swenson S, Yeh PJF, Wahr J, and Famiglietti J (2006) A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophysical Research Letters* 33: L16401.

Syed TH, Famiglietti JS, Chen J, et al. (2005) Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance. *Geophysical Research Letters* 32(24): L24404.

Syed TH, Famiglietti JS, Rodell M, Chen J, and Wilson CR (2008) Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research* 44: W02433. http://dx.doi.org/10.1029/2006WR005779.

Tamisiea ME and Davis JL (2006) Trends in the geoid over North America. *EOS Transactions American Geophysical Union* 87(36). Jt. Assem. Suppl., Abstract G32A-01.

Tamisiea ME, Leuliette EW, Davis JL, and Mitrovica JX (2005) Constraining hydrological and cryospheric mass flux in southeastern Alaska using space-based gravity measurements. *Geophysical Research Letters* 32: L20501.

Tapley BD, Bettadpur S, Ries JC, Thompson PF, and Watkins MM (2004a) GRACE measurements of mass variability in the Earth system. *Science* 305(5683): 503–505.

Tapley BD, Bettadpur S, Watkins M, and Reigber C (2004b) The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters* 31(9): L09607.

Thompson PF, Bettadpur SV, and Tapley BD (2004) Impact of short period, non-tidal, temporal mass variability on GRACE gravity estimates. *Geophysical Research Letters* 31(6): L06619.

Tiwari VM, Wahr J, and Swenson S (2009) Dwindling groundwater resources in northern India, from satellite gravity observations. *Geophysical Research Letters* 36: L18401. http://dx.doi.org/10.1029/2009GL039401.

Tregoning P, Watson C, Ramillien G, McQueen H, and Zhang J (2009) Detecting hydrologic deformation using GRACE and GPS. *Geophysical Research Letters* 36: L15401. http://dx.doi.org/10.1029/2009GL038718.

Trenberth KE, Smith L, Qian T, Dai A, and Fasullo J (2006) Estimates of the global water budget and its annual cycle using observational and model data. *Journal of Hydrometeorology* 8(4): 758–769. http://dx.doi.org/10.1175/JHM600.1.

Trupin AS, Meier MF, and Wahr JM (1992) The effect of melting glaciers on the Earth's rotation and gravitational field: 1965–1984. *Geophysical Journal International* 108: 1–15.

van Dam T, Wahr J, and Lavallee D (2007) A comparison of annual vertical crustal displacements from GPS and Gravity Recovery and Climate Experiment (GRACE) over Europe. *Journal of Geophysical Research* 112: B03404. http://dx.doi.org/10.1029/2006JB004335.

Velicogna I, Tong J, Zhang T, and Kimball JS (2012) Increasing subsurface water storage in discontinuous permafrost areas of the Lena River basin, Eurasia, detected from GRACE. *Geophysical Research Letters* 39: L09403. http://dx.doi.org/10.1029/2012GL051623.

Velicogna I and Wahr J (2002) A method for separating Antarctic postglacial rebound and ice mass balance using future ICESat Geoscience Laser Altimeter System, Gravity Recovery and Climate Experiment, and GPS satellite data. *Journal of Geophysical Research* 107(B10): 2263. http://dx.doi.org/10.1029/2001JB000708.

Velicogna I and Wahr J (2006a) Measurements of time variable gravity shows mass loss in Antarctica. *Science* 311: 1754–1756.

Velicogna I and Wahr J (2006b) Significant acceleration of Greenland ice mass loss in spring, 2004. *Nature*. http://dx.doi.org/10.1038/nature05168.

Velicogna I, Wahr J, and van den Dool H (2001) Can surface pressure be used to remove atmospheric contributions from GRACE data with sufficient accuracy to recover hydrological signals? *Journal of Geophysical Research* 106: 16415–16434.

Wahr JM and Bergen Z (1986) The effects of mantle anelasticity on nutations, Earth tides, and tidal variations in rotation rate. *Geophysical Journal of the Royal Astronomical Society* 87: 633–668.

Wahr J, Jayne SR, and Bryan FO (2002) A method of inferring deep ocean currents from satellite measurements of time variable gravity. *Journal of Geophysical Research: Oceans* 107: 3218. http://dx.doi.org/10.1029/2001JC001274.

Wahr J, Molenaar M, and Bryan F (1998) Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research* 103(B12): 30205–30229.

Wahr J, Swenson S, and Velicogna I (2006) The accuracy of GRACE mass estimates. *Geophysical Research Letters* 33: L06401. http://dx.doi.org/10.1029/2005GL025305.

Wang L, Shum CK, Simons FJ, Tapley B, and Dai C (2012) Coseismic and postseismic deformation of the 2011 Tohoku-Oki earthquake constrained by GRACE gravimetry. *Geophysical Research Letters* 39. http://dx.doi.org/10.1029/2012GL051104.

Watkins MM and Flechtner F (2012) Status of the GRACE follow-on mission. In: *2012 EGU General Assembly, Vienna*, p. 11403.

Willis JK, Chambers DP, and Nerem RS (2008) Assessing the globally averaged sea level budget on seasonal to interannual time scales. *Journal of Geophysical Research* 113: C06015. http://dx.doi.org/10.1029/2007JC004517.

Wouters B and Schrama EJO (2007) Improved accuracy of GRACE gravity solutions through empirical orthogonal function filtering of spherical harmonics. *Geophysical Research Letters* 34: L23711. http://dx.doi.org/10.1029/2007GL032098.

Yeh PJF, Swenson S, Famiglietti J, and Rodell M (2006) Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research* 42(12): W12203. http://dx.doi.org/10.1029/2006WR005374.

Yoder CF, Williams JG, Dickey JO, Schutz BE, Eanes RJ, and Tapley BD (1983) Secular variations of Earth's gravitational harmonic *Jsub*2 coefficient from Lageos and the non-tidal acceleration of Earth rotation. *Nature* 303: 757–762.

Zlotnicki V, Wahr J, and Song T (2007) Antarctic circumpolar current transport variability during 2003–2005 from GRACE. *Journal of Physical Oceanography* 37(2): 230–244.

This page intentionally left blank

3.09 Earth Rotation Variations – Long Period

RS Gross, California Institute of Technology, Pasadena, CA, USA

Published by Elsevier B.V.

3.09.1	**Introduction**	215
3.09.2	**Theory of Earth Rotation Variations at Long Periods**	216
3.09.2.1	Instantaneous Rotation Vector	216
3.09.2.2	Celestial Intermediate Pole	219
3.09.3	**Earth Rotation Measurement Techniques**	222
3.09.3.1	Lunar Occultation	222
3.09.3.2	Optical Astrometry	222
3.09.3.3	Space Geodesy	224
3.09.3.3.1	Very long-baseline interferometry	224
3.09.3.3.2	Global navigation satellite system	224
3.09.3.3.3	Satellite and lunar laser ranging	225
3.09.3.3.4	Doppler Orbitography and Radiopositioning Integrated by Satellite	225
3.09.3.4	Ring Laser Gyroscope	226
3.09.3.5	Intertechnique Combinations	226
3.09.4	**Observed and Modeled Earth Rotation Variations**	227
3.09.4.1	UT1 and LOD Variations	227
3.09.4.1.1	Secular trend, tidal dissipation, and glacial isostatic adjustment	228
3.09.4.1.2	Decadal variations and core–mantle interactions	230
3.09.4.1.3	Tidal variations and solid Earth, oceanic, and atmospheric tides	232
3.09.4.1.4	Seasonal variations	235
3.09.4.1.5	Interannual variations and ENSO	236
3.09.4.1.6	Intraseasonal variations and the Madden–Julian oscillation	237
3.09.4.2	Polar Motion	237
3.09.4.2.1	True polar wander and glacial isostatic adjustment	238
3.09.4.2.2	Decadal variations, the Markowitz wobble, and core–mantle interactions	240
3.09.4.2.3	Tidal wobbles and oceanic and atmospheric tides	241
3.09.4.2.4	The Chandler wobble and its excitation	243
3.09.4.2.5	Seasonal wobbles	244
3.09.4.2.6	Nonseasonal wobbles	246
3.09.5	**Mass Redistribution, Gravity, and Earth Rotation**	247
3.09.5.1	GRACE and SLR Measurements of Earth Rotation Excitation	248
Acknowledgments		248
References		248

3.09.1 Introduction

The Earth is a dynamic system – it has a fluid, mobile atmosphere and oceans; a continually changing global distribution of ice, snow, and water; a fluid core that is undergoing some type of hydromagnetic motion; a mantle both thermally convecting and rebounding from the glacial loading of the last ice age; and mobile tectonic plates. In addition, external forces due to the gravitational attraction of the Sun, Moon, and planets also act upon the Earth. These internal dynamical processes and external gravitational forces exert torques on the solid Earth or displace its mass, thereby causing the Earth's rotation to change.

Changes in the rotation of the solid Earth are studied by applying the principle of conservation of angular momentum to the Earth system. Under this principle, the rotation of the solid Earth changes as a result of (1) applied external torques, (2) internal mass redistribution, and (3) the transfer of angular momentum between the solid Earth and the fluid regions with which it is in contact; concomitant torques are due to hydrodynamic or magnetohydrodynamic stresses acting at the fluid/solid Earth interfaces.

Here, changes in the Earth's rotation that occur on timescales greater than a day are discussed. Using the principle of conservation of angular momentum, the equations governing small variations both in the rate of rotation and in the position of the rotation vector with respect to the Earth's crust are first derived. These equations are then rewritten in terms of the Earth rotation parameters that are actually reported by Earth rotation measurement services. The techniques that are used to monitor the Earth's rotation by the measurement services are then reviewed, a description of the variations that are observed by these techniques is given, and the possible causes of the observed variations are discussed.

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00059-2

3.09.2 Theory of Earth Rotation Variations at Long Periods

3.09.2.1 Instantaneous Rotation Vector

In a rotating reference frame that has been attached in some manner to the solid body of the Earth, the Euler equation of motion that relates changes in the angular momentum $\mathbf{L}(t)$ of the Earth to the external torques $\boldsymbol{\tau}(t)$ acting on it is (Eubanks, 1993; Lambeck, 1980, 1988; Moritz and Mueller, 1988; Munk and MacDonald, 1960; Seitz and Schuh, 2010)

$$\frac{\partial \mathbf{L}(t)}{\partial t} + \boldsymbol{\omega}(t) \times \mathbf{L}(t) = \boldsymbol{\tau}(t) \qquad [1]$$

where, strictly speaking, $\boldsymbol{\omega}(t)$ is the angular velocity of the rotating frame with respect to inertial space. But since the rotating frame has been attached to the solid body of the Earth, it is also interpreted as being the angular velocity of the Earth with respect to inertial space. In general, the angular momentum $\mathbf{L}(t)$ can be written as the sum of two terms: (1) that part $\mathbf{h}(t)$ due to motion relative to the rotating reference frame and (2) that part due to the changing inertia tensor $\mathbf{I}(t)$ of the Earth, which is changing because the distribution of the Earth's mass is changing:

$$\mathbf{L}(t) = \mathbf{h}(t) + \mathbf{I}(t) \cdot \boldsymbol{\omega}(t) \qquad [2]$$

Combining eqns [1] and [2] yields the Liouville equation

$$\frac{\partial}{\partial t}[\mathbf{h}(t) + \mathbf{I}(t) \cdot \boldsymbol{\omega}(t)] + \boldsymbol{\omega}(t) \times [\mathbf{h}(t) + \mathbf{I}(t) \cdot \boldsymbol{\omega}(t)] = \boldsymbol{\tau}(t) \qquad [3]$$

The external torques acting on the Earth due to the gravitational attraction of the Sun, Moon, and planets cause the Earth to nutate and precess. Since the nutations and precession of the Earth are discussed in **Chapter 3.10** of this volume, the external torques $\boldsymbol{\tau}(t)$ in eqn [3] will be set to zero. Note, however, that tidal effects on the Earth's rotation, which are also caused by the gravitational attraction of the Sun, Moon, and planets, are discussed here in **Sections 3.09.4.1.3** and **3.09.4.2.3**.

The Earth's rotation deviates only slightly from a state of uniform rotation, the deviation being a few parts in 10^8 in speed, corresponding to changes of a few milliseconds (ms) in the length of day, and about a part in 10^6 in the position of the rotation axis with respect to the crust of the Earth, corresponding to a variation of several hundred milliarcseconds (mas) in polar motion. Such small deviations in rotation can be studied by linearizing eqn [3]. Let the Earth initially be uniformly rotating at the constant rate Ω about the z-coordinate axis of the body-fixed reference frame and orient the frame within the Earth in such a manner that the inertia tensor of the Earth is diagonal in this frame:

$$\boldsymbol{\omega}_0 = \Omega \hat{\mathbf{z}} \qquad [4]$$

$$\mathrm{I}_0 = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} \qquad [5]$$

where the hat denotes a vector of unit length, Ω is the mean angular velocity of the Earth, and A, B, and C are the mean principal moments of inertia of the Earth ordered such that $A < B < C$. In this initial state, in which all time-dependent quantities vanish, the Earth is rotating at a constant rate about its figure axis, there are no mass displacements, and there is no relative angular momentum. So, for example, the atmosphere, oceans, and core are at rest with respect to the solid Earth and merely corotate with it.

Now let this initial state be perturbed by the appearance of mass displacements and relative angular momentum. In general, since the crust and mantle of the Earth can deform, they can undergo motion relative to the rotating reference frame and hence can contribute to the relative angular momentum. However, let the body-fixed reference frame in the perturbed state be oriented in such a manner that the relative angular momentum due to the motion of the crust and mantle vanishes. In this frame, which is known as the Tisserand mean-mantle frame of the Earth (Tisserand, 1891), the motion of the atmosphere, oceans, and core has a relative angular momentum, but the motion of the crust and mantle does not.

In the Tisserand mean-mantle frame, the perturbed instantaneous rotation vector and inertia tensor of the Earth can be written without loss of generality as

$$\boldsymbol{\omega}(t) = \boldsymbol{\omega}_\mathrm{o} + \Delta\boldsymbol{\omega}(t) = \Omega\hat{\mathbf{z}} + \Omega\left[m_x(t)\hat{\mathbf{x}} + m_y(t)\hat{\mathbf{y}} + m_z(t)\hat{\mathbf{z}}\right] \qquad [6]$$

$$\begin{aligned} \mathbf{I}(t) &= \mathbf{I}_\mathbf{o} + \Delta\mathbf{I}(t) \\ &= \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} + \begin{pmatrix} \Delta I_{xx}(t) & \Delta I_{xy}(t) & \Delta I_{xz}(t) \\ \Delta I_{xy}(t) & \Delta I_{yy}(t) & \Delta I_{yz}(t) \\ \Delta I_{xz}(t) & \Delta I_{yz}(t) & \Delta I_{zz}(t) \end{pmatrix} \end{aligned} \qquad [7]$$

where the terms with the subscript o denote the initial values given by eqns [4] and [5], the $\Omega m_i(t)$ are the elements of the time-dependent perturbation $\Delta\boldsymbol{\omega}(t)$ to the rotation vector, and the $\Delta I_{ij}(t)$ are the elements of the time-dependent perturbation $\Delta\mathbf{I}(t)$ to the inertia tensor.

The equation that relates small changes in the Earth's rotation to the mass displacements and relative angular momenta that are causing the rotation to change can be derived by substituting eqns [6] and [7] into eqn [3] and then linearizing the resulting expression by assuming that the $h_i(t) \ll \Omega C$, the $m_i(t) \ll 1$, and the $\Delta I_{ij}(t) \ll C$. By keeping terms to first order in these small quantities, the equatorial and axial components of eqn [3] can be written as

$$\frac{1}{\sigma_r}\frac{\partial m_x(t)}{\partial t} + \left[\frac{B(C-B)}{A(C-A)}\right]^{1/2} m_y(t) = -\left(\frac{B}{A}\right)^{1/2}\left[\frac{1}{\Omega}\frac{\partial \phi_{r,x}(t)}{\partial t} - \phi_{r,y}(t)\right] \qquad [8]$$

$$\frac{1}{\sigma_r}\frac{\partial m_y(t)}{\partial t} - \left[\frac{A(C-A)}{B(C-B)}\right]^{1/2} m_x(t) = -\left(\frac{A}{B}\right)^{1/2}\left[\frac{1}{\Omega}\frac{\partial \phi_{r,y}(t)}{\partial t} + \phi_{r,x}(t)\right] \qquad [9]$$

$$\frac{1}{\Omega}\frac{\partial m_z(t)}{\partial t} = -\frac{1}{\Omega}\frac{\partial \phi_{r,z}(t)}{\partial t} \qquad [10]$$

where the external torques have been set to zero,

$$\sigma_r^2 = \left(\frac{C-A}{A}\right)\left(\frac{C-B}{B}\right)\Omega^2 \qquad [11]$$

and the $\phi_{r,i}(t)$, known as excitation functions, are

$$\phi_{r,x}(t) = \frac{h_x(t) + \Omega\Delta I_{xz}(t)}{\Omega\sqrt{(C-A)(C-B)}} \qquad [12]$$

$$\phi_{r,y}(t) = \frac{h_y(t) + \Omega \Delta I_{yz}(t)}{\Omega\sqrt{(C-A)(C-B)}} \quad [13]$$

$$\phi_{r,z}(t) = \frac{1}{C\Omega}[h_z(t) + \Omega \Delta I_{zz}(t)] \quad [14]$$

Equations [8] and [9] are coupled, first-order differential equations that describe the motion of the rotation pole in the rotating, body-fixed reference frame as it responds to the applied excitation. In the absence of excitation, the solution of these equations, which describes the natural or free motion of the rotation pole, can be written as

$$m_x(t) = m\cos(\sigma_r t + \alpha) \quad [15]$$

$$m_y(t) = \left[\frac{A(C-A)}{B(C-B)}\right]^{1/2} m\sin(\sigma_r t + \alpha) \quad [16]$$

where m is the amplitude of the motion along the x-axis and α is the phase of the motion. The natural motion described by eqns [15] and [16] is prograde undamped elliptical motion of frequency σ_r. Using the values in **Table 1** for A, B, and C of the whole Earth, the period of the natural frequency, given by $2\pi/\sigma_r$, is found to be 304.46 sidereal days.

Table 1 Geodetic parameters of the Earth

Parameter	*Value*	*Source*
G	6.67259×10^{-11} m^3 (kg s^2)$^{-1}$	(a)
M_{atm}	5.1441×10^{18} kg	(b)
M_{ocn}	1.4×10^{21} kg	(c)
Whole Earth (*observed*)		
Ω	7.292115×10^{-5} rad s^{-1}	(a)
M	5.9737×10^{24} kg	(a)
C	8.0365×10^{37} kg m^2	(a)
B	8.0103×10^{37} kg m^2	(a)
A	8.0101×10^{37} kg m^2	(a)
$C-A$	2.6398×10^{35} kg m^2	(a)
$C-B$	2.6221×10^{35} kg m^2	(a)
$B-A$	1.765×10^{33} kg m^2	(a)
Whole Earth (*modeled*)		
a	6371.0 km	(d)
n_0	0.15505	(e)
k_2	0.298	(f)
$\Delta k_{\mathrm{ocn,w}}$	0.047715	(g)
$\Delta k_{\mathrm{ocn,s}}$	0.043228	(g)
k_r	0.997191	(g)
k'_2	-0.305	(f)
$\Delta k'_{\mathrm{an}}$	$-0.011 + i\,0.003$	(f)
α_3	0.792	(h)
Crust and mantle (*PREM*)		
ε_a	3.334×10^{-3}	(d)
M_m	4.0337×10^{24} kg	(d)
C_m	7.1236×10^{37} kg m^2	(d)
A_m	7.0999×10^{37} kg m^2	(d)
Core (*PREM*)		
ε_c	2.546×10^{-3}	(d)
M_c	1.9395×10^{24} kg	(d)
C_c	9.1401×10^{36} kg m^2	(d)
A_c	9.1168×10^{36} kg m^2	(d)

PREM, Preliminary Reference Earth Model (Dziewonski and Anderson, 1981).
(a) Groten (2004), (b) Trenberth and Guillemot (1994), (c) Yoder (1995), (d) Mathews et al. (1991), (e) Dahlen (1976), (f) Wahr (2005), (g) this work, and (h) Wahr (1983).

Euler (1765) first predicted that the Earth should freely wobble as it rotates and that the period of this free wobble, assuming that the Earth is rigid, would be about 10 months. However, it was not until 1891 that the free wobble of the Earth was first detected in astronomical observations by Seth Carlo Chandler, Jr. (Chandler, 1891), albeit at a period of 14 months. The free wobble of the Earth is now known as the Chandler wobble in his honor.

Equations [8]–[14] describe changes in the rotation of a rigid body of arbitrary shape that is subject to small perturbing excitation. By recognizing that $(B-A)/A = 2.2 \times 10^{-5} \ll 1$ for the Earth (see **Table 1**), so that dynamically the Earth is nearly axisymmetric, eqns [8] and [9] can be simplified by replacing A and B in them with the average $A' = (A+B)/2$ of the equatorial principal moments of inertia of the Earth:

$$\frac{1}{\sigma_{ra}}\frac{\partial m_x(t)}{\partial t} + m_y(t) = \phi_{ra,y}(t) - \frac{1}{\Omega}\frac{\partial \phi_{ra,x}(t)}{\partial t} \quad [17]$$

$$\frac{1}{\sigma_{ra}}\frac{\partial m_y(t)}{\partial t} - m_x(t) = -\phi_{ra,x}(t) - \frac{1}{\Omega}\frac{\partial \phi_{ra,y}(t)}{\partial t} \quad [18]$$

where the excitation functions $\phi_{ra,i}(t)$ of a rigid axisymmetric body are

$$\phi_{ra,x}(t) = \frac{h_x(t) + \Omega \Delta I_{xz}(t)}{(C-A')\Omega} \quad [19]$$

$$\phi_{ra,y}(t) = \frac{h_y(t) + \Omega \Delta I_{yz}(t)}{(C-A')\Omega} \quad [20]$$

In the absence of excitation, the free motion of a rigid axisymmetric body is prograde undamped circular motion of natural frequency

$$\sigma_{ra} = \left(\frac{C-A'}{A'}\right)\Omega \quad [21]$$

Note that eqns [10] and [14], which describe changes in the rate of rotation of a rigid body, are the same whether the body is dynamically axisymmetric or triaxial.

Equations [10], [14], and [17]–[21] describe changes in the rotation of a rigid axisymmetric body that is subject to small perturbing excitation. But the Earth is not rigid – it has an atmosphere and oceans, a fluid core, and a solid crust and mantle that can deform in response not only to the applied excitation but also to changes in rotation that are caused by the excitation. In general, changes in rotation can be expected to cause changes both in the Earth's inertia tensor and in relative angular momentum. However, by definition of the Tisserand mean-mantle frame, there are no changes in relative angular momentum caused by the motion of the crust and mantle. Furthermore, if it is assumed that the oceans stay in equilibrium as the rotation of the solid Earth changes so that no oceanic currents are generated by the changes in rotation, then there are also no changes in relative angular momentum due to the motion of the oceans. And the effects of the atmosphere can be ignored here because of its relatively small mass (see **Table 1**). Thus, only the core will contribute to changes in relative angular momentum caused by changes in rotation.

Smith and Dahlen (1981) used the results of Hough (1895) to show that the change $\delta h_i(\sigma)$ in relative angular momentum

due to core motion caused by a rigid rotation of an axisymmetric crust and mantle is

$$\begin{pmatrix}\delta h_x(\sigma)\\ \delta h_y(\sigma)\\ \delta h_z(\sigma)\end{pmatrix}=\begin{pmatrix}E & iE' & 0\\ -iE' & E & 0\\ 0 & 0 & \widetilde{E}\end{pmatrix}\begin{pmatrix}m_x(\sigma)\\ m_y(\sigma)\\ m_z(\sigma)\end{pmatrix} \quad [22]$$

where to first order in the ellipticity ε_c of the surface of the core and at frequencies $\sigma \ll \Omega$

$$E=(\sigma^2/\Omega)A_c \quad [23]$$

$$E'=-\sigma(1-\varepsilon_c)A_c \quad [24]$$

$$\widetilde{E}=-\Omega C_c \quad [25]$$

where A_c and C_c are the equatorial and axial principal moments of inertia of the core, respectively, and eqn [25] for $\widetilde{E}$ has been inferred here by realizing that the core cannot respond to axial changes in the rotation of the mantle if the core–mantle boundary is axisymmetric and if there is no coupling between the core and the mantle (Merriam, 1980; Wahr et al., 1981; Yoder et al., 1981). Note that because of the assumption of dynamical axisymmetry, the equatorial components of eqn [22] are uncoupled from the axial, so that no spin–wobble coupling is introduced by the response of the core to changes in rotation.

Dahlen (1976) studied the passive influence of the oceans on the Earth's rotation, including the changes δI_{ij} in the Earth's inertia tensor that are caused by changes in the rotation of the Earth. In the absence of oceans and assuming that the Earth responds to the centripetal potential associated with changes in rotation in exactly the same manner that a nonrotating Earth would respond to a static potential of the same amplitude and type, Dahlen (1976) found

$$\begin{pmatrix}\delta I_{xz}\\ \delta I_{yz}\\ \delta I_{zz}\end{pmatrix}=\frac{a^5\Omega^2}{3G}\begin{pmatrix}k_2 & 0 & 0\\ 0 & k_2 & 0\\ 0 & 0 & n_o+\frac{4}{3}k_2\end{pmatrix}\begin{pmatrix}m_x\\ m_y\\ m_z\end{pmatrix} \quad [26]$$

where k_2 is the second-degree body tide Love number of the whole Earth (not just of the mantle (see Smith and Dahlen, 1981, p. 239; although for a different opinion, see Dickman, 2005), n_o arises from the change in the mean moment of inertia of the Earth caused by the term in the centripetal potential that gives rise to a purely radial deformation of the Earth, G is the Newtonian gravitational constant, and a is the mean radius of the Earth (i.e., the radius of a sphere having the same volume as the Earth).

When equilibrium oceans are present, Dahlen (1976) found that eqn [26] for the changes in the inertia tensor caused by changes in rotation is modified to

$$\begin{pmatrix}\delta I_{xz}\\ \delta I_{yz}\\ \delta I_{zz}\end{pmatrix}=\begin{pmatrix}D+\delta D & \delta D_{12} & \delta D_{13}\\ \delta D_{12} & D-\delta D & \delta D_{23}\\ \delta D_{13} & \delta D_{23} & \widetilde{D}\end{pmatrix}\begin{pmatrix}m_x\\ m_y\\ m_z\end{pmatrix} \quad [27]$$

where

$$D=(k_2+\Delta k_{ocn,w})\frac{a^5\Omega^2}{3G} \quad [28]$$

$$\widetilde{D}=\left[n_o+\tfrac{4}{3}(k_2+\Delta k_{ocn,s})\right]\frac{a^5\Omega^2}{3G} \quad [29]$$

where the influence of equilibrium oceans has been written in terms of an 'oceanic Love number' Δk_{ocn}, which modifies the second-degree body tide Love number k_2. Dahlen (1976) found that because of the nonuniform distribution of the oceans, the oceanic Love number is different for each component. However, the average of the equatorial components has been taken here to define a mean oceanic Love number $\Delta k_{ocn,w}$ for the wobble. From eqn [27], it is seen that the nonuniform distribution of the oceans has also coupled the equatorial components to the axial via the off-diagonal elements δD_{13} and δD_{23}. However, the numerical results of Dahlen (1976) for the Earth model 1066A (Gilbert and Dziewonski, 1975) show that this coupling is very weak, with $\delta D_{13}/D=2.17\times10^{-3}$ and $-\delta D_{23}/D=0.55\times10^{-3}$. The coupling between the equatorial components is also very weak, with $-\delta D_{12}/D=3.15\times10^{-3}$.

Using eqns [22]–[25] to account for the relative angular momentum of the core caused by changes in rotation, using eqns [27]–[29] to account for both the rotational deformation of the Earth and the passive response of equilibrium oceans to changes in rotation, and keeping terms to first order in small quantities, the linearized Liouville equation becomes

$$\frac{1}{\sigma_{cw}}\frac{\partial m_x(t)}{\partial t}+m_y(t)=\chi_y(t)-\frac{1}{\Omega}\frac{\partial \chi_x(t)}{\partial t} \quad [30]$$

$$\frac{1}{\sigma_{cw}}\frac{\partial m_y(t)}{\partial t}-m_x(t)=-\chi_x(t)-\frac{1}{\Omega}\frac{\partial \chi_y(t)}{\partial t} \quad [31]$$

$$\frac{1}{\Omega}\frac{\partial m_z(t)}{\partial t}=-\frac{1}{\Omega}\frac{\partial \chi_z(t)}{\partial t} \quad [32]$$

where the theoretical frequency of the Chandler wobble is

$$\sigma_{cw}=\left(\frac{C-A'-D}{A'_m+\varepsilon_c A_c+D}\right)\Omega \quad [33]$$

where $A'_m=A'-A_c$ is the equatorial principal moment of inertia of the crust and mantle, and the excitation functions $\chi_i(t)$ are

$$\chi_x(t)=\frac{h_x(t)+\Omega(1+k'_2)\Delta I_{xz}(t)}{(C-A'-D)\Omega} \quad [34]$$

$$\chi_y(t)=\frac{h_y(t)+\Omega(1+k'_2)\Delta I_{yz}(t)}{(C-A'-D)\Omega} \quad [35]$$

$$\chi_z(t)=k_r\frac{h_z(t)+\Omega(1+\alpha_3 k'_2)\Delta I_{zz}(t)}{C_m\Omega} \quad [36]$$

where C_m is the axial principal moment of inertia of the crust and mantle and k_r is a factor, whose value is near unity (see **Table 1**), that accounts for the effects of rotational deformation on the axial component:

$$k_r=\left\{1+\left[n_o+\tfrac{4}{3}(k_2+\Delta k_{ocn,s})\right]\frac{a^5\Omega^2}{3G}\frac{1}{C_m}\right\}^{-1} \quad [37]$$

The deformation of the Earth associated with surficial excitation processes that load the solid Earth has been taken into account in eqns [34]–[36] by including the second-degree load Love number k'_2 where, because of core decoupling, the load Love number in the axial component is modified by a factor of α_3 (Dickman, 2003; Merriam, 1980; Nam and Dickman, 1990; Wahr, 1983). Expressions for the excitation functions for processes that do not load the solid Earth can be recovered from eqns [34]–[36] by setting the load Love number k'_2 to zero.

Equations [30]–[36] describe changes in the rotation of an elastic axisymmetric body having a fluid core and equilibrium oceans that is subject to small perturbing excitation. Equation [33] for the theoretical Chandler frequency of such a body was first derived by Smith and Dahlen (1981). Applying this result to the Earth, they found that the elasticity of the solid Earth lengthens the period of the Chandler wobble from the rigid Earth value by 143.0 sidereal days, deformation of the oceans lengthens it a further 29.8 sidereal days, and the presence of a fluid core decreases it by 50.5 sidereal days. Using the values in **Table 1** for the Earth, the theoretical period of the Chandler wobble is found to be 426.8 sidereal days, or about 7.4 sidereal days shorter than the observed period of 434.2 ± 1.1 (1σ) sidereal days (Wilson and Vicente, 1990). This discrepancy between the theoretical and observed periods of the Chandler wobble is probably mainly due to the effects of mantle anelasticity, since departures of the oceans from equilibrium as large as 1% increase the Chandler period by only 0.3 sidereal days (Smith and Dahlen, 1981).

Mantle anelasticity modifies the body tide Love number k_2 and hence, via D, the frequency of the Chandler wobble and the equatorial excitation functions. It also modifies the load Love number k'_2. In the absence of accurate models of mantle anelasticity at the frequencies of interest here, namely, at frequencies $\sigma < \Omega$, a hybrid approach is taken to include its effects. The observed complex-valued frequency σ_o of the Chandler wobble is substituted for its theoretical value in eqns [30] and [31]. It is also substituted for its theoretical value in the excitation functions after rewriting them in terms of the theoretical value by using eqn [33] to eliminate D. The results are

$$\frac{1}{\sigma_o}\frac{\partial m_x(t)}{\partial t} + m_y(t) = \chi_y(t) - \frac{1}{\Omega}\frac{\partial \chi_x(t)}{\partial t} \qquad [38]$$

$$\frac{1}{\sigma_o}\frac{\partial m_y(t)}{\partial t} - m_x(t) = -\chi_x(t) - \frac{1}{\Omega}\frac{\partial \chi_y(t)}{\partial t} \qquad [39]$$

$$m_z(t) = -\chi_z(t) \qquad [40]$$

where the excitation functions become

$$\chi_x(t) = \frac{h_x(t) + \Omega\left[1 + \left(k'_2 + \Delta k'_{\mathrm{an}}\right)\right]\Delta I_{xz}(t)}{\left[C - A' + A'_m + \varepsilon_c A_c\right]\sigma_o} \qquad [41]$$

$$\chi_y(t) = \frac{h_y(t) + \Omega\left[1 + \left(k'_2 + \Delta k'_{\mathrm{an}}\right)\right]\Delta I_{yz}(t)}{\left[C - A' + A'_m + \varepsilon_c A_c\right]\sigma_o} \qquad [42]$$

$$\chi_z(t) = k_r \frac{h_z(t) + \Omega\left[1 + \alpha_3\left(k'_2 + \Delta k'_{an}\right)\right]\Delta I_{zz}(t)}{C_m \Omega} \qquad [43]$$

where $\Delta k'_{\mathrm{an}}$ accounts for the effects of mantle anelasticity on the load Love number. Equations [38]–[43] are the final expressions for the changes $m_i(t)$ in the rotation of the Earth caused by small excitation $\chi_i(t)$. Numerically, using 434.2 sidereal days (Wilson and Vicente, 1990) for the observed period of the Chandler wobble and the values in **Table 1** for the other constants, the real parts of the excitation functions can be written as

$$\chi_x(t) = \frac{1.608\left[h_x(t) + 0.684\,\Omega\Delta I_{xz}(t)\right]}{(C - A')\Omega} \qquad [44]$$

$$\chi_y(t) = \frac{1.608\left[h_y(t) + 0.684\,\Omega\Delta I_{yz}(t)\right]}{(C - A')\Omega} \qquad [45]$$

$$\chi_z(t) = \frac{0.997}{C_m \Omega}\left[h_z(t) + 0.750\,\Omega\Delta I_{zz}(t)\right] \qquad [46]$$

These results agree with those of Wahr (1982, 1983, 2005) to within 2%, with most of the disagreement being due to differences in the values of the numerical constants.

The approach used here to derive the linearized Liouville equations that can be used to study small changes in the Earth's rotation follows the approach of Smith and Dahlen (1981) and Wahr (1982, 1983, 2005). Other approaches have been given by Barnes et al. (1983), Eubanks (1993) (also see Aoyama and Naito, 2000), Dickman (1993, 2003), and Jochmann (2009). Dickman (2003) compared some of these different approaches and discussed the implications of nonzero coupling between the core and mantle; Wahr (2005) discussed the implications of mantle anelasticity. A frequency-dependent transfer function approach has been used by Chen and Shen (2010) to develop a theory of the Earth's rotation that accounts for triaxiality of the mantle and core, anelasticity of the mantle, and dissipation in the oceans. The influence of triaxiality on oceanless elastic bodies with a fluid core, with application to the rotation of Mars, has been studied by Yoder and Standish (1997) and Van Hoolst and Dehant (2002).

3.09.2.2 Celestial Intermediate Pole

Small changes in the Earth's rotation caused by small changes in relative angular momentum or small changes in the Earth's inertia tensor can be studied using eqns [3]–[43] where $\Omega m_i(t)$ are the elements of the change $\Delta\boldsymbol{\omega}(t)$ to the Earth rotation vector, so that $m_x(t)$, $m_y(t)$, and $1 + m_z(t)$ are the direction cosines of the rotation vector with respect to the coordinate axes of the rotating, body-fixed terrestrial reference frame. Alternatively, $m_x(t)$ and $m_y(t)$ can be interpreted as being the angular offsets of the rotation vector from the $\hat{\mathbf{z}}$-axis of the rotating reference frame in the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ directions. That is, $m_x(t)$ and $m_y(t)$ specify the location of the rotation pole within the rotating, body-fixed terrestrial reference frame, where the rotation pole is that point defined by the intersection of the rotation axis with the surface of the Earth near the North Pole. But Earth rotation measurement services do not report the location of the rotation pole within the rotating, body-fixed terrestrial reference frame. Instead, they report the location of the celestial intermediate pole (CIP).

Just three time-dependent angles, the Euler angles, are required to directly transform the coordinates of some station from the terrestrial frame to the celestial frame. These angles are time-dependent, of course, because the Earth, and hence the body-fixed terrestrial reference frame, is rotating. But, by tradition, an intermediate reference frame is used with the result that five angles are required to completely transform station coordinates from the terrestrial to the celestial reference frames (e.g., Sovers et al., 1998):

$$\mathbf{r}_c(t) = \mathbf{PNSXY}\,\mathbf{r}_t(t) \qquad [47]$$

where $\mathbf{r}_t(t)$ are the, in general time-dependent, coordinates of the station in the rotating, body-fixed terrestrial frame; $\mathbf{r}_c(t)$ are

the coordinates of the station in the celestial frame; and **P**, **N**, **S**, **X**, and **Y** are the classical transformation matrices with P accounting for the precession of the Earth, **N** accounting for nutation, **S** accounting for spin, and **X** and **Y** accounting for the x- and y-components of polar motion. By first applying **X** and **Y**, the terrestrial coordinates of the station are transformed to an intermediate frame whose reference pole is the CIP; **S** represents a spin through a large angle about the $\hat{z}$-axis of the intermediate frame; **P** and **N** finally transform the intermediate frame to the celestial frame. This approach of using an intermediate frame and five angles (two polar motion parameters, two nutation parameters, and a spin parameter) to transform station coordinates between the terrestrial and celestial reference frames has been traditionally followed in order to separate polar motion from precession and nutation. This separation is done in such a manner that the precessional and nutational motion of the Earth is long period when observed in the celestial reference frame and polar motion is long period when observed in the terrestrial reference frame.

Earth rotation measurement services report the parameters that are needed to carry out the transformation given by eqn [47], namely, the polar motion parameters $p_x(t)$ and $p_y(t)$ that are required in **X** and **Y** and that give the location of the CIP in the rotating, body-fixed terrestrial reference frame, the nutation parameters $\delta\psi(t)$ and $\delta\varepsilon(t)$ that are required in **N** and that are corrections in longitude and obliquity to the adopted nutation model that are needed to give the location of the CIP in the celestial reference frame, and a spin parameter UT1(t) that is required in **S** and that represents the angle through which the Earth has rotated. The precession transformation matrix **P** depends on the lunisolar and planetary precession constants.

The relationship between the polar motion parameters $p_x(t)$ and $p_y(t)$ that are reported by Earth rotation measurement services and the elements $\Omega m_x(t)$ and $\Omega m_y(t)$ of the Earth rotation vector that are needed in eqns [38] and [39] can be derived by considering the properties of transformation matrices (Goldstein, 1950). The transformation of station coordinates between two frames having a common origin implies a rotation. Applying the transformation matrix to the position vector of some station to get its coordinates in a new frame is equivalent to a rotation of the coordinate axes. If the initial reference frame is the terrestrial reference frame of the Earth and the final frame is the celestial reference frame and because the terrestrial reference frame has been fixed to the body of the Earth, then the equivalent rotation of the coordinate axes is simply the rotation of the Earth. Because an intermediate frame has been used to separate polar motion from precession and nutation, in order to derive the relationship between the polar motion parameters $p_x(t)$ and $p_y(t)$ and the elements $\Omega m_x(t)$ and $\Omega m_y(t)$ of the Earth rotation vector, it is sufficient to consider the transformation matrix $\mathbf{A}^{\mathrm{T}}=\mathbf{SXY}$ that transforms station coordinates between the terrestrial and intermediate reference frames:

$$\mathbf{r}_i(t)=\mathbf{A}^{\mathrm{T}}(t)\mathbf{r}_t(t) \qquad [48]$$

where the superscript T denotes the transpose and $\mathbf{r}_i(t)$ are the coordinates of the station in the intermediate frame. The elements ω_i of the rotation vector that is associated with this transformation matrix, that is, of the rotation vector of the Earth, are the three independent elements of the antisymmetric matrix $\mathbf{W}(t)$ (Gross, 1992; Kinoshita et al., 1979):

$$\mathbf{W}(t)=\frac{\mathrm{d}\mathbf{A}(t)}{\mathrm{d}t}\mathbf{A}^{\mathrm{T}}(t)=\begin{pmatrix}0 & \omega_z & -\omega_y\\ -\omega_z & 0 & \omega_x\\ \omega_y & -\omega_x & 0\end{pmatrix} \qquad [49]$$

From Sovers et al. (1998), the classical **X**, **Y**, and **S** transformation matrices are

$$\mathbf{X}(t)=\begin{pmatrix}\cos p_x(t) & 0 & -\sin p_x(t)\\ 0 & 1 & 0\\ \sin p_x(t) & 0 & \cos p_x(t)\end{pmatrix} \qquad [50]$$

$$\mathbf{Y}(t)=\begin{pmatrix}1 & 0 & 0\\ 0 & \cos p_y(t) & \sin p_y(t)\\ 0 & -\sin p_y(t) & \cos p_y(t)\end{pmatrix} \qquad [51]$$

$$\mathbf{S}(t)=\begin{pmatrix}\cos H(t) & -\sin H(t) & 0\\ \sin H(t) & \cos H(t) & 0\\ 0 & 0 & 1\end{pmatrix} \qquad [52]$$

where by tradition the positive direction of $p_y(t)$ is taken to be toward 90°W longitude and H is the hour angle of the true equinox of date that is related to UT1 and the hour angle of the mean equinox of date by the equation of the equinoxes. By forming $\mathbf{A}^{\mathrm{T}}=\mathbf{SXY}$, using eqn [49], and keeping terms to first order in small quantities, the desired elements of the rotation vector of the terrestrial frame with respect to the intermediate frame is obtained (Brzezinski, 1992; Brzezinski and Capitaine, 1993; Gross, 1992):

$$\omega_x(t)=\Omega p_x(t)-\frac{\mathrm{d}p_y(t)}{\mathrm{d}t} \qquad [53]$$

$$\omega_y(t)=-\Omega p_y(t)-\frac{\mathrm{d}p_x(t)}{\mathrm{d}t} \qquad [54]$$

$$\omega_z(t)=[1+p_z(t)]\Omega \qquad [55]$$

where the time rate of change of H has been set equal to $(1+p_z)\ \Omega$ where $p_z(t)=m_z(t)$ represents small departures from uniform spin at the mean sidereal rotation rate Ω of the Earth.

In complex notation, with $\mathbf{m}(t)=m_x(t)+im_y(t)$ and $\mathbf{p}(t)=p_x(t)-ip_y(t)$, where the negative sign accounts for $p_y(t)$ being positive toward 90°W longitude, eqns [53] and [54] can be written as

$$\mathbf{m}(t)=\mathbf{p}(t)-\frac{i}{\Omega}\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} \qquad [56]$$

For frequencies of motion $\sigma \ll \Omega$, the second term on the right-hand side of eqn [56] becomes much smaller than the first and the motion of the rotation pole becomes the same as the motion of the CIP. But for frequencies of motion $|\sigma| \approx \Omega$, the motions of the rotation and CIPs are very different. This difference becomes important when studying rapid motions such as those caused by the diurnal and semidiurnal ocean tides (see **Section 3.09.4.2.3**). For example, the amplitude of the prograde diurnal tide-induced motion of the rotation pole is about twice as large as that of the CIP.

Using eqns [53] and [54] for the relationship between the reported polar motion parameters $p_x(t)$ and $p_y(t)$ and the elements $\omega_x(t)$ and $\omega_y(t)$ of the Earth rotation vector, eqns [38] and [39] can be written in complex notation as (Brzezinski, 1992; Gross, 1992)

$$\mathbf{p}(t) + \frac{i}{\sigma_o}\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = \boldsymbol{\chi}(t) \quad [57]$$

where $\boldsymbol{\chi}(t) = \boldsymbol{\chi}_x(t) + i\chi_y(t)$. The axial component, eqn [40], is usually written in terms of changes $\Delta\Lambda(t)$ of the length of day as

$$\frac{\Delta\Lambda(t)}{\Lambda_o} = \chi_z(t) \quad [58]$$

where Λ_o is the nominal length of day (LOD) of 86 400 s. Equations [57] and [58] are the final expressions, written in terms of the parameters actually reported by Earth rotation measurement services, for the changes in the rotation of the Earth caused by small excitation $\chi_i(t)$, where the excitation functions are given by eqns [41]–[43].

Since five angles are traditionally used to transform station coordinates between the terrestrial and celestial reference frames when only three are required, the five traditional Earth Orientation Parameters (EOPs) are not independent of each other. Because the frequency σ_c of some motion as observed in the celestial reference frame is related to the frequency σ_t of that same motion as observed in the terrestrial reference frame by

$$\sigma_c = \sigma_t + \Omega \quad [59]$$

then motion having a retrograde nearly diurnal frequency in the terrestrial reference frame ($\sigma_t \approx -\Omega$) will be of low frequency (long period) when observed in the celestial reference frame. That is, retrograde nearly diurnal polar motions are equivalent to nutations. In particular, the two polar motion parameters are related to the two nutation parameters by (Brzezinski, 1992; Brzezinski and Capitaine, 1993)

$$\mathbf{p}(t) = -\mathbf{n}(t)\mathrm{e}^{-\mathrm{i}\Omega t} \quad [60]$$

where Greenwich Mean Sidereal Time (GMST) has been approximated by Ωt in the exponent and $\mathbf{n}(t) = \delta\psi(t)\sin\varepsilon_o + i\,\delta\varepsilon(t)$ with ε_o being the mean obliquity of the ecliptic.

A degeneracy also exists between the EOPs and different realizations of the terrestrial reference frame. Since by eqn [49] the Earth rotation vector can be determined from the transformation matrix that transforms station coordinates between the terrestrial and celestial reference frames, if the realization of the terrestrial reference frame changes, then the elements of the rotation vector will change. In particular, a positive change in the x-component of polar motion is equivalent to a left-handed (clockwise) rotation of the terrestrial reference frame about the $\hat{\mathbf{y}}$-axis; a positive change in the y-component of polar motion, remembering that $p_y(t)$ is defined to be positive toward 90°W longitude, is equivalent to a left-handed rotation of the terrestrial reference frame about the $\hat{\mathbf{x}}$-axis; and a positive change in UT1 is equivalent to a right-handed (counterclockwise) rotation of the terrestrial reference frame about the $\hat{\mathbf{z}}$-axis.

The polar motion parameters $p_x(t)$ and $p_y(t)$ give the location in the terrestrial frame of the reference pole of the intermediate frame, whatever that intermediate frame may be. The 1980 International Astronomical Union (IAU) theory of nutation adopted the celestial ephemeris pole (CEP) as the reference pole of the intermediate frame (Seidelmann, 1982), defining it to be a pole that exhibits no nearly diurnal motions either in the body-fixed terrestrial frame or in the celestial frame. The CEP was chosen (Seidelmann, 1982) to be the B-axis of Wahr (1981), which is the axis of figure for the Tisserand mean outer surface of the Earth, where the averaging procedure is such that the resulting B-axis does not move in response to body tides. Since observing stations are attached to the outer surface of the Earth, their measurements are sensitive to the motion of the Earth's surface in space. Wahr (1981) thus generalized the concept of the Tisserand mean-mantle frame to that of the Tisserand mean surface frame, with his B-axis being the reference axis of the Tisserand mean surface frame that moves in space with the mean motion of the observing stations.

In 2000, the IAU adopted the CIP as the reference pole of the intermediate frame (Petit and Luzum, 2010, **Chapter 1.06**). The definition of the CIP extends that of the CEP by clarifying the definition of polar motion and precession–nutation. The CEP was defined in such a manner that it exhibits no nearly diurnal motion in either the terrestrial or celestial reference frames. That is, precession–nutation were considered to be the motion of the CEP as viewed in the celestial reference frame with the frequency of motion ranging between −0.5 and +0.5 cycles per sidereal day (cpsd), and polar motion was considered to be the motion of the CEP as viewed in the terrestrial reference frame with the frequency of motion in that frame ranging between −0.5 and +0.5 cpsd (Capitaine, 2000). Since frequencies of motion in the two frames are related by eqn [59], in the celestial reference frame, polar motion was the motion of the CEP with frequencies ranging between +0.5 and +1.5 cpsd. Thus, the motion of the CEP in the celestial frame was defined for frequencies between −0.5 and +1.5 cpsd, with the division between polar motion and precession–nutation being at a frequency of +0.5 cpsd. The motion of the CEP outside this celestial frequency band was undefined. Similarly, the motion of the CEP in the terrestrial reference frame was defined for frequencies between −1.5 and +0.5 cpsd with the division between polar motion and precession–nutation being at a frequency of −0.5 cpsd. The motion of the CEP outside this terrestrial frequency band was also undefined (see **Figure 1**).

Since 1980, when the CEP was adopted as the reference pole of the intermediate frame, models of polar motion with frequencies outside the terrestrial frequency band within which the motion of the CEP was defined became available. These polar motions were due to the effects of diurnal and semidiurnal ocean tides. Models of nutations having frequencies outside the celestial frequency band within which the CEP was defined also became available, as did space-geodetic measurements having subdaily temporal resolution. With these improvements in models and measurements came the need to extend the definition of the CEP to all possible frequencies of motion, not only those between −0.5 and +1.5 cpsd in the celestial frame (−1.5 and +0.5 cpsd in the terrestrial frame). At the time that the definition of the intermediate pole was extended, its name was changed to the CIP. When defining the CIP, the concept used in defining the CEP of a pole having no nearly diurnal motions in either the terrestrial or celestial reference frames had to be abandoned because ocean tides can cause polar motions having frequencies near +1 cpsd as viewed in the terrestrial reference frame.

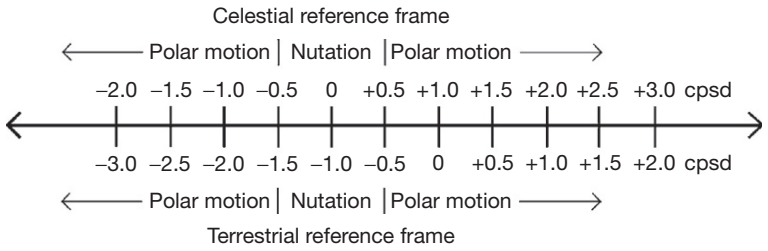

Celestial reference frame
Polar motion
Nutation
Polar motion
−2.0 −1.5 −1.0 −0.5 0 +0.5 +1.0 +1.5 +2.0 +2.5 +3.0 cpsd
−3.0 −2.5 −2.0 −1.5 −1.0 −0.5 0 +0.5 +1.0 +1.5 +2.0 cpsd
Polar motion
Nutation
Polar motion
Terrestrial reference frame

Figure 1 Schematic illustration of the relationship between the frequency of some motion as viewed in the celestial (top half of figure) and terrestrial (bottom half of figure) reference frames (see eqn [59]). By convention, nutation is the motion of the CIP having frequencies in the range between −0.5 and +0.5 cpsd as viewed in the celestial frame. As viewed in the terrestrial frame, this same nutational motion of the Earth has frequencies in the range between −1.5 and −0.5 cpsd. Motion at all other frequencies is considered to be polar motion.

The CIP is still chosen to be the axis of figure for the Tisserand mean outer surface of the Earth, as it was for the CEP. And the CIP is defined in such a manner that precession and nutation are still considered to be the motion of the CIP as viewed in the celestial reference frame with the frequency of motion ranging between −0.5 and +0.5 cpsd (Capitaine, 2000). But now polar motion is considered to be the motion of the CIP in the celestial frame at all other frequencies. That is, polar motion is considered to be the motion of the CIP in the terrestrial frame at all frequencies except those between −1.5 and −0.5 cpsd (see **Figure 1**). This has the effect of including in nutation those ocean tidal terms having retrograde nearly diurnal frequencies as viewed in the terrestrial reference frame and of including in polar motion those nutation terms having frequencies less than −0.5 cpsd or greater than +0.5 cpsd as viewed in the celestial reference frame (see **Table 2**).

3.09.3 Earth Rotation Measurement Techniques

Changes in the Earth's rate of rotation become apparent when comparing time kept by the rotating Earth, known as Universal Time (UT), to uniform timescales based either upon atomic clocks or upon the motion of the Sun and other celestial bodies. Prior to the development of atomic clocks, the most accurate measurements of changes in the Earth's rate of rotation were obtained by timing the occultations of stars by the Moon. With the advent of atomic clocks in 1955, a uniform atomic timescale became available that could be used as a reference when measuring the transit times of stars as they pass through the local meridian. Changes in the Earth's rate of rotation could then be determined more accurately from optical astrometric measurements of star transits than they could from measurements of lunar occultations. And prior to the development of space-geodetic techniques, optical astrometric measurements of changes in the apparent latitudes of observing stations yielded the most accurate estimates of polar motion. The space-geodetic techniques of very long-baseline interferometry (VLBI), global navigation satellite systems (GNSSs) like the Global Positioning System (GPS), and satellite laser ranging and lunar laser ranging (SLR and LLR, respectively) are now the most accurate techniques available for measuring changes both in the Earth's rate of rotation and in polar motion.

3.09.3.1 Lunar Occultation

The most recent rereduction of lunar occultation measurements for UT and LOD changes is that of Jordi et al. (1994) who analyzed about 53 000 observations of lunar occultations spanning 1830.0–1955.5. They used a reference frame defined by the FK5 star catalog, the LE200 lunar ephemeris, and corrections for the limb profile of the Moon. The UT series they obtained consists of values and 1σ uncertainties for the difference between Terrestrial Time (TT) and UT form 1 (TT–UT1) spanning 1830.0–1955.5 at 4-month intervals. TT is a dynamical timescale that can be related to the International Atomic Time (TAI) by adding 32.184 s to TAI (McCarthy and Guinot, 2012; McCarthy and Seidelmann, 2009; Seidelmann et al., 1992).

Jordi et al. (1994) extended their (TT–UT1) series to 1992 by using values of (UT1–TAI) obtained from the Bureau International de l'Heure (BIH) and the International Earth Rotation and Reference Systems Service (IERS). They then derived an LOD series spanning 1830–1987 at 4-month intervals by finite differencing and smoothing the extended UT1 series. Gross (2001) combined the lunar occultation measurements of Jordi et al. (1994) with optical astrometric and space-geodetic measurements to produce a smoothed LOD series spanning 1832.5–1997.5 at yearly intervals. Other UT1 and LOD series based upon lunar occultation measurements are those of Morrison (1979), Stephenson and Morrison (1984, 1985), McCarthy and Babcock (1986), and Liao and Greiner-Mai (1999).

3.09.3.2 Optical Astrometry

The International Latitude Service (ILS) was established by the International Association of Geodesy (IAG) in 1895 (e.g., Höpfner, 2000a) for the purpose of monitoring the wobbling motion of the Earth that had been detected by Seth Carlo Chandler, Jr., in 1891. As the Earth wobbles, the apparent latitude of an astronomical observing station will vary. To measure this variation of latitude and infer the underlying polar motion that is causing it, the ILS established six observing stations that were well distributed in longitude and that were all located at nearly the same latitude of 39°8′N. A seventh station, Kitab, was added in 1930 to replace the station at Tschardjui that ceased operations in 1919 due to a nearby river changing its course and adversely affecting the

Table 2 Coefficients of those nutation terms that are included in polar motion by the definition of the CIP

	Fundamental argument							$p_x(t)$ *(μas)*		$p_y(t)$ *(μas)*	
Degree of potential	γ	l	l′	F	D	Ω	*Period (solar days)*	*sin*	*cos*	*sin*	*cos*
2	1	1	0	0	0	0	0.96244	0.76	−0.43	0.43	0.76
2	1	0	0	0	0	−1	0.99712	1.93	−1.11	1.11	1.93
2	1	0	0	0	0	0	0.99727	14.27	−8.19	8.19	14.27
2	1	0	0	−2	2	−2	1.00275	−4.76	2.73	−2.73	−4.76
2	1	−1	0	0	0	0	1.03472	0.84	−0.48	0.48	0.84
2	1	0	0	−2	0	−2	1.07581	−11.36	6.52	−6.52	−11.36
2	1	0	0	−2	0	−1	1.07598	−2.14	1.23	−1.23	−2.14
2	1	1	0	−2	−2	−2	1.22346	−0.44	0.25	−0.25	−0.44
2	1	−1	0	−2	0	−2	1.11951	−2.31	1.32	−1.32	−2.31
2	1	−1	0	−2	0	−1	1.11970	−0.44	0.25	−0.25	−0.44
3	0	1	0	1	0	1	13.719	1.28	0.16	−0.16	1.28
3	0	0	0	1	0	0	27.212	2.62	0.32	−0.32	2.62
3	0	0	0	1	0	1	27.322	16.64	2.04	−2.04	16.64
3	0	0	0	1	0	2	27.432	−0.87	−0.11	0.11	−0.87
3	0	1	0	1	−2	1	193.560	2.10	0.27	−0.27	2.10
3	0	0	0	1	−1	1	365.242	1.31	0.20	−0.20	1.31
3	0	1	1	−1	0	−1	411.807	1.05	0.27	−0.27	1.05
3	0	1	1	−1	0	0	438.360	−0.63	0.12	−0.12	−0.63
3	0	−1	0	1	0	0	2190.35	−2.78	−0.31	0.31	−2.78
3	0	−1	0	1	0	1	3231.50	−16.16	−1.83	1.83	−16.16
3	0	−1	0	1	0	2	6159.14	0.78	0.09	−0.09	0.78
3	0	1	0	−1	0	−2	−6159.14	−0.68	−0.09	0.09	−0.68
3	0	1	0	−1	0	−1	−3231.50	12.32	1.59	−1.59	12.32
3	0	1	0	−1	0	0	−2190.35	1.86	0.24	−0.24	1.86
3	0	−1	0	−1	2	−1	−193.560	0.81	0.10	−0.10	0.81
3	0	0	0	−1	0	−2	−27.432	−0.82	−0.10	0.10	−0.82
3	0	0	0	−1	0	−1	−27.322	15.75	1.93	−1.93	15.75
3	0	0	0	−1	0	0	−27.212	2.48	0.30	−0.30	2.48
3	0	−1	0	−1	0	−1	−13.719	1.39	0.17	−0.17	1.39
Rate of secular polar motion (μas/year) due to the zero frequency tide											
4	0	0	0	0	0	0			−3.80		−4.31

Terms with amplitudes less than 0.5 microarcseconds (μas) are not tabulated. γ is GMST reckoned from the lower culmination of the vernal equinox (GMST + π). l, l', F, D, and Ω are the Delaunay arguments, expressions for which are given in Simon et al. (1994). The period, given in solar days, is the approximate period of the term as viewed in the terrestrial reference frame. Terms having positive (negative) periods indicate prograde (retrograde) circular motion. Summing prograde and retrograde circular motions having the same period yields elliptical motion. The nearly diurnal terms, like the nearly diurnal and nearly semidiurnal ocean tidal terms, are not included in the polar motion parameters reported by Earth rotation measurement services. However, the secular rate and the long-period terms, like the long-period ocean tidal terms, are included in the reported polar motion parameters.

Source: Mathews PM and Bretagnon P (2003) Polar motions equivalent to high frequency nutations for a nonrigid Earth with anelastic mantle. *Astronomy and Astrophysics* 400: 1113–1128, with permission from Springer. Also, see Petit and Luzum (2010, Table 5.1a.)

seeing conditions at Tschardjui. Locating all the ILS stations at nearly the same latitude allowed common star pairs to be observed by the same Horrebow–Talcott method (Munk and MacDonald, 1960, **Chapter 1.09**), thereby allowing the polar motion to be determined from the latitude observations free of first-order errors in the reference star catalog.

The use of different star catalogs, standards, and data reduction procedures during the history of the ILS observing program can introduce discontinuities in the polar motion series derived from the ILS observations. In order to produce a homogeneous polar motion series unaffected by these sources of error, Yumi and Yokoyama (1980) rereduced 772 395 latitude observations taken at the seven ILS observing stations using the Melchior and Dejaiffe (1969) star catalog and the 1964 IAU System of Astronomical Constants. The resulting polar motion series, known as the homogeneous ILS series, spans from October 1899 to December 1978 at monthly intervals.

During the twentieth century, numerous other optical astrometric measurements of latitude and longitude were taken at other stations and by other methods besides those of the ILS. At the BIH, Li (1985) and Li and Feissel (1986) rereduced 240 140 optical astrometric measurements of longitude and 259 159 measurements of latitude taken at 136 observing stations. In order to produce an Earth orientation series independent of the ILS series, no measurements taken at the ILS stations were used. The resulting UT1 and polar motion series, which is known as the BIH series, spans from 5 January 1962 to 31 December 1981 at 5-day intervals.

The High-Precision Parallax Collecting Satellite (Hipparcos) star catalog, being constructed from observations taken above the Earth's atmosphere, is substantially more accurate than earlier ground-based catalogs such as the Melchior and Dejaiffe (1969) catalog used by Yumi and Yokoyama (1980). The improved accuracy of the Hipparcos star catalog motivated

Vondrák (1991, 1999) and Vondrák et al. (1992, 1995, 1997, 1998) to once again rereduce optical astrometric measurements for Earth orientation parameters. All available optical astrometric measurements, numbering 4 315 628 from 48 instruments including those taken at the ILS stations, were collected and corrected for instrumental errors and such systematic effects as plate tectonic motion, ocean loading, and tidal variations. The corrected measurements were then used to estimate nutation, polar motion, and UT. The resulting Earth orientation series, known as the Hipparcos series, consists of values and uncertainties for polar motion and nutation spanning 1899.7–1992.0 at quasi-5-day intervals, with UT1 estimates starting in 1956 shortly after atomic clocks first became available.

3.09.3.3 Space Geodesy

An integral part of geodesy has always been the definition and realization of a terrestrial, body-fixed reference frame, a celestial, space-fixed reference frame, and the determination of the EOPs (precession, nutation, spin, and polar motion) that link these two reference frames together. But with the advent of space geodesy – with the placement of laser retroreflectors on the Moon by Apollo astronauts and Soviet landers, the launch of the Laser Geodynamics Satellite (LAGEOS), the development of VLBI, and the development of GNSSs like GPS – a quantum leap has been taken in our ability to realize the terrestrial and celestial reference frames and to determine the Earth Orientation Parameters.

The only space-geodetic measurement technique capable of independently determining all of the EOPs is multibaseline VLBI (see Chapter 3.11). All of the other techniques either need to apply external constraints to the determined EOPs or can determine only subsets of the EOPs, only linear combinations of the EOPs, or only their time rates of change.

3.09.3.3.1 Very long-baseline interferometry

Radio interferometry is routinely used to make highly accurate measurements of UT1 and polar motion with observing sessions lasting from about an hour to a day. The VLBI technique measures the difference in the arrival time of a radio signal at two or more radio telescopes that are simultaneously observing the same distant extragalactic radio source (Lambeck, 1988, Chapter 1.08; Robertson, 1991; Sovers et al., 1998). This technique is therefore sensitive to processes that change the relative position of the radio telescopes with respect to the source, such as a change in the orientation of the Earth in space or a change in the position of the telescopes due to, for example, tidal displacements or tectonic motions. If just two telescopes are observing the same source, then only two components of the Earth's orientation can be determined. A rotation of the Earth about an axis parallel to the baseline connecting the two radio telescopes does not change the relative position of the telescopes with respect to the source, and hence, this component of the Earth's orientation is not determinable from VLBI observations taken on that single baseline. Multibaseline VLBI observations with satisfactory geometry can determine all of the components of the Earth's orientation including their time rates of change.

The International VLBI Service for Geodesy and Astrometry (IVS; Schlüter et al., 2002; Schlüter and Behrend, 2007), a service of both the IAG and the IAU, was established on 11 February 1999 to support research in geodesy, geophysics, and astrometry. As part of its activities, it coordinates the acquisition and reduction of VLBI observations for the purpose, in part, of monitoring changes in the Earth's rotation and defining and maintaining the international terrestrial and celestial reference frames. VLBI data products, including EOPs determined from both single and multibaseline observations, are available through the IVS website (see Table 3).

Table 3 Sources of Earth orientation data

Data type	*URL*
Very long-baseline interferometry	
IVS	http://ivscc.gsfc.nasa.gov/
Global navigation satellite system	
IGS	http://igscb.jpl.nasa.gov/
Satellite and lunar laser ranging	
ILRS	http://ilrs.gsfc.nasa.gov/
DORIS	
IDS	http://ids.cls.fr/
Intertechnique combinations	
IERS	http://www.iers.org/
KEOF	http://keof.jpl.nasa.gov/

URL, uniform resource locator; IVS, International VLBI Service for Geodesy and Astrometry; VLBI, very long-baseline interferometry; IGS, International GNSS Service; GNSS, global navigation satellite system; ILRS, International Laser Ranging Service; DORIS, Doppler Orbitography and Radiopositioning Integrated by Satellite; IDS, International DORIS Service; IERS, International Earth Rotation and Reference Systems Service; KEOF, Kalman Earth Orientation Filter.

3.09.3.3.2 Global navigation satellite system

GNSSs consist of two major elements: (1) a space-based element consisting of a constellation of transmitting satellites and (2) a ground-based element consisting of a network of receivers. In the United States' GPS, the satellites, including spares, are at altitudes of 20 200 km in orbits with periods of 11 h 58 min located in six orbital planes each inclined at 55° to the Earth's equator with four or more satellites in each plane. In the Global Navigation Satellite System (GLONASS) of Russia, the satellites are at altitudes of 19 100 km in circular orbits with periods of 11 h 15 min located in three orbital planes each inclined at 64.8° to the Earth's equator with eight satellites in each plane. In the Galileo system of Europe, which is not yet fully operational, the satellites will be at altitudes of 23 222 km in circular orbits with periods of 14 h 22 min located in three orbital planes each inclined at 56° to the Earth's equator with nine satellites and one spare in each plane.

In GNSSs, the navigation signals are broadcast by the satellites at more than one frequency, thereby enabling first-order corrections to be made for ionospheric refraction effects. The ground-based multichannel receivers detect the navigation signals being broadcast by those satellites that are above the horizon (up to the number of channels in the receiver). In principle, trilateration can then be used to determine the position of each receiver and, by extension, the orientation of the network of receivers as a whole. In practice, in order to achieve higher accuracy, more sophisticated analysis techniques are employed to determine the EOPs and other quantities such as orbital parameters of the satellites, positions of the stations,

and atmospheric parameters such as the zenith path delay (Beutler et al., 1996; Blewitt, 1993, 2007; Bock and Leppard, 1990; Hofmann-Wellenhof et al., 2008; Leick, 2003).

Only polar motion and its time rate of change can be independently determined from GNSS measurements. UT1 cannot be separated from the orbital elements of the satellites and hence cannot be determined from GNSS data. The time rate of change of UT1, which is related to the length of day, can be determined from GNSS measurements. But because of the corrupting influence of orbit error, VLBI measurements are usually used to constrain the GNSS-derived LOD estimates.

The International GNSS Service (IGS; Beutler et al., 1999, 2009; Dow et al., 2005, 2009), a service of the IAG, was established on 1 January 1994 under its former name of the International GPS Service to support Earth science research. As part of its activities, it coordinates the acquisition and reduction of GNSS observations for the purpose, in part, of maintaining the International Terrestrial Reference Frame (ITRF) and monitoring changes in the Earth's rotation and geocenter motion. GNSS data products, including EOPs, are available through the IGS website (see **Table 3**).

3.09.3.3.3 Satellite and lunar laser ranging

In the technique of SLR, the round-trip flight times of laser light pulses are accurately measured as they are emitted from a laser system located at some ground-based observing station, travel through the Earth's atmosphere to some artificial satellite orbiting the Earth, are reflected by retroreflectors carried onboard that satellite, and return to the same observing station from which they were emitted (Combrinck, 2010; Exertier et al., 2006; Lambeck, 1988, **Chapter 1.07**). This time-of-flight range measurement is converted into a distance measurement by using the speed of light and correcting for a variety of known or modeled effects such as atmospheric path delay and satellite center-of-mass offset. Although a number of satellites carry retroreflectors for tracking and navigation purposes, the LAGEOS I and II satellites were specifically designed and launched to study geodetic properties of the Earth including its rotation and are the satellites most commonly used to determine the EOPs. Including range measurements to the Etalon I and II satellites has been found to strengthen the solution for the EOPs, so these satellites are now often included when determining the EOPs.

The EOPs are recovered from the basic range measurements in the course of determining the satellite's orbit. The basic range measurement is sensitive to any geophysical process that changes the distance between the satellite and the observing station, such as displacements of the satellite due to perturbations of the Earth's gravitational field, motions of the observing station due to tidal displacements or plate tectonics, or a change in the orientation of the Earth (which changes the location of the observing station with respect to the satellite). These and other geophysical processes must be modeled when fitting the satellite's orbit to the range measurements as obtained at a number of globally distributed tracking stations. Adjustments to the a priori models used for these effects can then be obtained during the orbit determination procedure, thereby enabling, for example, the determination of station positions and EOPs (Smith et al., 1985, 1990, 1994; Tapley et al., 1985, 1993). However, because variations in UT1 cannot be separated from variations in the orbital node of the satellite, which are caused by the effects of unmodeled forces acting on the satellite, it is not possible to independently determine UT1 from SLR measurements. But independent estimates of the time rate of change of UT1, or equivalently of LOD, can be determined from SLR measurements, as can polar motion and its time rate of change.

The technique of LLR is similar to that of SLR except that the laser retroreflector is located on the Moon instead of on an artificial satellite (Dickey et al., 1994a; Exertier et al., 2006; Lambeck, 1988, **Chapter 1.09**; Mulholland, 1980; Shelus, 2001; Williams et al., 1993, 2009). LLR is technically more challenging than SLR because of the need to detect the much weaker signal that is returned from the Moon. Larger, more powerful laser systems with more sophisticated signal detection systems need to be employed in LLR; consequently, there are far fewer stations that range to the Moon than range to artificial satellites. In fact, there are currently only three stations that are active in ranging to the Moon: the Observatoire de la Côte d'Azur in the south of France, the McDonald Observatory in Texas, and the Apache Point Observatory in New Mexico.

The EOPs are typically determined from LLR data by analyzing the residuals at each station after the lunar orbit and other parameters such as station and reflector locations have been fit to the range measurements (Dickey et al., 1985; Langley et al., 1981a; Stolz et al., 1976). From this single-station technique, two linear combinations of UT1 and the polar motion parameters $p_x(t)$ and $p_y(t)$ can be determined, namely, UT0 and the variation of latitude $\Delta\phi_i(t)$, at that station:

$$\Delta\phi_i(t) = p_x(t)\cos\lambda_i - p_y(t)\sin\lambda_i \qquad [61]$$

$$\mathbf{UT0}(\mathrm{t}) - \mathbf{TAI}(\mathrm{t}) = \mathbf{UT1}(\mathrm{t}) - \mathbf{TAI}(\mathrm{t}) + \left[p_x(t)\sin\lambda_\mathrm{i} + p_y(t)\cos\lambda_\mathrm{i}\right]\tan\phi_\mathrm{i} \qquad [62]$$

where ϕ_i and λ_i are the nominal latitude and longitude of station i. A rotation of the Earth about an axis connecting the station with the origin of the terrestrial reference frame does not change the distance between the station and the Moon, and hence, this component of the Earth's orientation cannot be determined from single-station LLR observations.

The International Laser Ranging Service (ILRS; Gurtner et al., 2005; Pearlman et al., 2002, 2005) is a service of the IAG that was established on 22 September 1998 to support research in geodesy, geophysics, and lunar science. As part of its activities, the ILRS coordinates the acquisition and reduction of SLR and LLR observations for the purpose, in part, of maintaining the ITRF and monitoring the Earth's rotation and geocenter motion. SLR and LLR data products, including EOPs, are available through the ILRS website (see **Table 3**).

3.09.3.3.4 Doppler Orbitography and Radiopositioning Integrated by Satellite

Like GNSSs, the French Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) system also consists of space-based and ground-based elements (Tavernier et al., 2005, 2006; Willis et al., 2006, 2010). But, in the DORIS system, the transmitting beacons are located on the ground and the receivers are located on the satellites. Currently, there are about 60 globally distributed beacons broadcasting to

receivers onboard six satellites (Jason-1, OSTM/Jason-2, CryoSat-2, HY-2A, SPOT-4, and SPOT-5). In the past, the SPOT-2, SPOT-3, TOPEX/Poseidon, and Envisat satellites carried DORIS receivers. And current plans call for the future SARAL, Sentinel-3A, Sentinel-3B, and Jason-3 satellites to carry DORIS receivers.

In the DORIS system, the Doppler shift of two transmitted frequencies is accurately measured. The use of two frequencies allows corrections to be made for ionospheric effects. Processing these Doppler measurements, usually as range-rate (Tapley et al., 2004a), allows the orbit of the satellite to be determined along with other quantities such as station positions and EOPs. As with other satellite techniques, UT1 cannot be determined from DORIS measurements, but its time rate of change can be determined, as can polar motion and its rate of change (Gambis, 2006; Willis et al., 2006).

The International DORIS Service (IDS; Tavernier et al., 2005, 2006; Willis et al., 2010) is a service of the IAG that was established on 1 July 2003 to support research in geodesy and geophysics. As part of its activities, the IDS coordinates the acquisition and reduction of DORIS observations for the purpose, in part, of maintaining the international terrestrial reference frame and monitoring the Earth's rotation and geocenter motion. DORIS data products, including EOPs, are available through the IDS website (see **Table 3**).

3.09.3.4 Ring Laser Gyroscope

Ring laser gyroscopes are a promising emerging technology for determining the Earth's rotation. In a ring laser gyroscope, two laser beams propagate in opposite directions around a ring. Since the ring laser gyroscope is rotating with the Earth, the effective path length of the beam that is corotating with the Earth is slightly longer than the path that is counterrotating with it. Because the effective path lengths of the two beams differ, their frequencies differ, so they interfere with each other to produce a beat pattern (Stedman, 1997). The beat frequency $\delta f(t)$, which can be measured and is known as the Sagnac frequency, is proportional to the instantaneous angular velocity $\boldsymbol{\omega}(t)$ of the Earth:

$$\delta f(t) = \frac{4A}{\lambda P}\mathbf{n}\cdot\boldsymbol{\omega}(t) \qquad [63]$$

where A is the area of the ring, P is the optical path length, λ is the wavelength of the laser beam in the absence of rotation, and $\mathbf{n}$ is the unit vector normal to the plane of the ring. By making the instrument rigid and the laser stable so that A, P, and λ are constants, the component of the Earth's instantaneous rotation vector that is parallel to the normal of the plane of the ring can be determined from measurements of the Sagnac frequency. All three components of the Earth rotation vector are determinable from three mutually orthogonal colocated ring laser gyroscopes or from a globally distributed network of gyroscopes.

Ring laser gyroscopes measure the absolute rotation of the Earth in the sense that, in principle, just a single measurement of the Sagnac frequency is required to determine $\mathbf{n}\cdot\boldsymbol{\omega}$. All of the other techniques discussed in the preceding text are relative sensors because they infer the Earth's rotation from the change in the orientation of the Earth that takes place between at least two measurements that are separated in time. Thus, ring laser gyroscopes are fundamentally different from the space-geodetic techniques that are currently used to regularly monitor the Earth's rotation. Measurements taken by ring laser gyroscopes can therefore be expected to contribute to our understanding of the Earth's changing rotation, on subdaily to daily timescales (Mendes Cerveira et al., 2009; Nilsson et al., 2012; Schreiber et al., 2004, 2009) and on longer timescales (Schreiber et al., 2011).

3.09.3.5 Intertechnique Combinations

EOPs can be determined from measurements taken by each of the techniques discussed in the preceding text. But each technique has its own unique strengths and weaknesses in this regard. Not only is each technique sensitive to a different subset and/or linear combination of the EOPs, but also the averaging time for their determination is different, as is the interval between observations, the precision with which they can be determined, and the duration of the resulting EOP series. By combining the individual series determined by each technique, a series of the Earth's orientation can be obtained that is based upon independent measurements and that spans the greatest possible time interval. Such a combined Earth orientation series is useful for a number of purposes, including a variety of scientific studies and as an a priori series for use in data reduction procedures. However, care must be taken when generating such a combined series to account for differences in the underlying reference frames within which each individual series is determined (which can lead to differences in the bias and rate of the Earth orientation series). Care must also be taken to properly assign weights to the observations prior to combination.

The IERS, a service of both the IAG and the IAU, was established on 1 January 1988 under its former name of the International Earth Rotation Service to serve the astronomical, geophysical, and geodetic communities by, in part, providing the International Celestial Reference Frame (ICRF), the ITRF, and the EOPs that are needed to transform station coordinates between these two frames (Dick and Richter, 2004; Feissel and Gambis, 1993; Vondrák and Richter, 2004). As part of its activities, the IERS combines and predicts EOPs determined by the space-geodetic techniques using a weighted-average approach (Bizouard and Gambis, 2009; Gambis, 2004; Gambis and Luzum, 2011; Johnson et al., 2005; Luzum et al., 2001). The EOPs produced by the IERS are available through its website (see **Table 3**).

Since the 1980s, a Kalman filter, the Kalman Earth Orientation Filter (KEOF), has been used at the Jet Propulsion Laboratory (JPL) to combine and predict EOPs in support of interplanetary spacecraft tracking and navigation (Freedman et al., 1994a; Gross et al., 1998). A Kalman filter has many properties that make it an attractive method of combining Earth orientation series. It allows the full accuracy of the measurements to be used, whether they are degenerate or full rank, are irregularly or regularly spaced in time, or are corrupted by systematic or other errors that can be described by stochastic models. And by using a stochastic model for the process, a Kalman filter can objectively account for the growth in the uncertainty of the EOPs between measurements. The

combined and predicted Earth orientation series produced at JPL using a Kalman filter are available through the KEOF website (see Table 3).

3.09.4 Observed and Modeled Earth Rotation Variations

The Earth's rotation changes on all observable timescales, from subdaily to decadal and longer. The wide range of timescales on which the Earth's rotation changes reflects the wide variety of processes that are causing it to change, including external tidal forces; surficial fluid processes involving the atmosphere, oceans, and hydrosphere; and internal processes acting both within the solid Earth itself and between the fluid core and solid Earth. The changes in the Earth's rotation that are observed and the geophysical models that have been developed to explain them are reviewed here. Sources of Earth rotation data are given in Table 3 and sources of geophysical models are given in Table 4.

Two general approaches are used to investigate why and how the Earth's rotation changes: the torque approach and the angular momentum approach (Munk and MacDonald, 1960; Wahr, 1982). In the torque approach, changes in the Earth's rotation are studied by computing the torques acting on the Earth's crust and mantle by, for example, processes due to the overlying atmosphere and oceans (de Viron and Dehant, 1999; de Viron et al., 1999, 2001a,b; Marcus et al., 2004). In the angular momentum approach, changes in the Earth's rotation are studied by invoking the principle of the conservation of angular momentum. Under this principle, if the angular momentum of, say, the atmosphere changes, then the angular momentum of the crust and mantle must change by an equal but opposite amount in order for the total angular momentum of the solid Earth–atmosphere system to remain the same.

The advantage of the torque approach is that it can be used to elucidate the mechanism by which angular momentum is exchanged between, say, the atmosphere and the solid Earth. In general, there are three different types of torques that act to exchange angular momentum between a surface geophysical fluid like the atmosphere and the solid Earth: (1) a pressure or mountain torque due to normal pressure forces acting on topographic features of the solid Earth, (2) a friction torque due to tangential frictional stresses caused by relative motion between a fluid like the atmosphere and the solid Earth, and (3) a gravitational torque due to gravitational attraction between density heterogeneities in the fluid and solid Earth. For example, Salstein and Rosen (1994) showed that mountain torques, and in particular those over South America, were largely responsible for the rapid exchange of axial angular momentum between the atmosphere and solid Earth that occurred at the beginning of August 1992.

The disadvantage of the torque approach is that the currently available atmospheric, oceanic, and hydrologic models are not suited for the computation of torques. Many parameters of the models, such as friction drag, that are needed to compute the torques are not known very well. And the grid scales of the models are too coarse to allow accurate computation of the torques (de Viron and Dehant, 2003). However, the models are well suited for the computation of angular momentum. So the angular momentum approach is most often used to investigate the causes of the observed changes in the Earth's rotation.

Table 4 Sources of angular momentum models

Model type	*URL*
Global geophysical fluids	
IERS GGFC	http://geophy.uni.lu/
Atmospheric	
IERS SBA	http://www.aer.com/science-research/earth/earth-mass-and-rotation/special-bureau-atmosphere
TU Vienna	http://ggosatm.hg.tuwien.ac.at/
Oceanic	
IERS SBO	http://euler.jpl.nasa.gov/sbo/
Hydrologic	
IERS SBH	http://www.csr.utexas.edu/research/ggfc/
Consistent AAM, *OAM*, and *HAM*	
GFZ Potsdam	http://www.gfz-potsdam.de/portal/gfz/Struktur/Departments/Department+1/sec13/services

URL, uniform resource locator; IERS, International Earth Rotation and Reference Systems Service; GGFC, Global Geophysical Fluids Center; SBA, Special Bureau for the Atmosphere; TU, Technical University; SBO, Special Bureau for the Oceans; SBH, Special Bureau for Hydrology; AAM, atmospheric angular momentum; OAM, ocean angular momentum; HAM, hydrological angular momentum; GFZ, GeoForschungsZentrum.

3.09.4.1 UT1 and LOD Variations

Observations of the length of day (Figure 2) show that it consists mainly of (1) a linear trend of rate +1.8 ms cy^{-1}, (2) decadal variations having an amplitude of a few milliseconds, (3) tidal variations having an amplitude of about 1 ms, (4) seasonal variations having an amplitude of about 0.5 ms, and

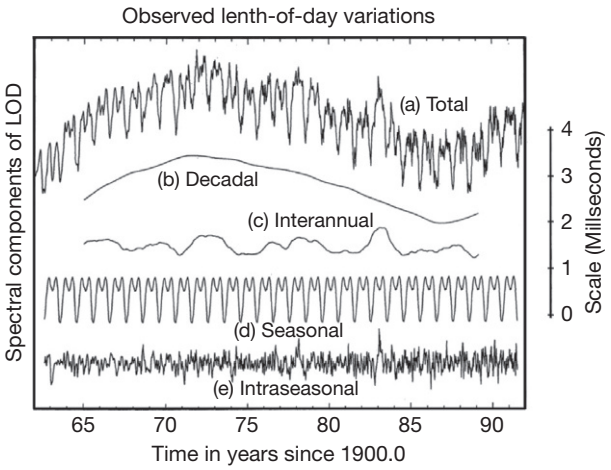


Figure 2 (a) Observed LOD variations during 1962–91 from the COMB91 combined optical astrometric and space-geodetic Earth orientation series and its decomposition into variations on (b) decadal, (c) interannual, (d) seasonal, and (e) intraseasonal timescales. Tidal variations are not shown. Reproduced from Dickey JO, Marcus SL, Hide R, Eubanks TM, and Boggs DH (1994b) Angular momentum exchange among the solid Earth, atmosphere and oceans: A case study of the 1982–83 El Niño event. *Journal of Geophysical Research* 99(B12): 23921–23937.

(5) smaller amplitude variations occurring on all measurable timescales. Here, the LOD variations that are observed and the models that have been developed to explain them are reviewed.

The length of day is the rotational period of the Earth. Changes $\Delta\Lambda(t)$ in the length of day are related to changes in UT and to changes $\Delta\omega_z(t) = \Omega m_z(t)$ in the axial component of the Earth's angular velocity by

$$\frac{\Delta\Lambda(t)}{\Lambda_o} = -m_z(t) = -\frac{\mathrm{d}(\mathrm{UT1} - \mathrm{TAI})}{\mathrm{d}t} \qquad [64]$$

By eqns [43] and [58], the length of day will change as a result of changes in the axial component $h_z(t)$ of relative angular momentum and of changes in the axial component $\Omega\Delta I_{zz}(t)$ of angular momentum due to changes in the mass distribution of the Earth. Modeling the observed changes in LOD therefore requires computing both types of changes in angular momentum for the different components of the Earth system. By definition, the angular momentum vector $\mathbf{L}(t)$ of some component of the Earth system such as the atmosphere or oceans is given by

$$\mathbf{L}(t) = \int_{V(t)} \rho(\mathbf{r}, t)\mathbf{r} \times [\boldsymbol{\omega} \times \mathbf{r} + \mathbf{u}(\mathbf{r}, t)]\,\mathrm{d}V \qquad [65]$$

where the integral is taken over the, in general, time-dependent volume $V(t)$ of that component of the Earth system under consideration, the first term in the square brackets represents changes in the angular momentum due to changes in mass distribution, the second term represents changes due to relative motion, and $\mathbf{r}$ is the position vector of some mass element of density $\rho(\mathbf{r},t)$ that is moving with the Eulerian velocity $\mathbf{u}(\mathbf{r},t)$ with respect to a terrestrial reference frame that (1) is fixed to the solid Earth; (2) is oriented such that its x- and y-coordinate axes are along the Greenwich and 90°E meridians, respectively, with both axes lying in the equatorial plane; (3) has an origin located at the center of the Earth; and (4) is rotating with angular velocity $\boldsymbol{\omega}$ with respect to an inertial reference frame. In general, $\boldsymbol{\omega}$ is not constant in time but exhibits variations in both magnitude (related to changes in LOD) and direction (related to polar motion). However, these changes are small, and for the purpose of deriving expressions for the angular momentum of different components of the Earth system, it can be assumed that the terrestrial reference frame is uniformly rotating at the rate Ω about the z-coordinate axis: $\boldsymbol{\omega} = \Omega\hat{\mathbf{z}}$. In this case, the axial component of the mass term of the angular momentum can be written as

$$\Omega\Delta I_{zz}(t) = \Omega\int_{V(t)} \rho(\mathbf{r}, t)r^2\cos^2\phi\,\mathrm{d}V \qquad [66]$$

where ϕ is the north latitude. Similarly, the axial component of the motion term of the angular momentum can be written as

$$h_z(t) = \int_{V(t)} \rho(\mathbf{r}, t)r\cos\phi\, u(\mathbf{r}, t)\,\mathrm{d}V \qquad [67]$$

where $u(\mathbf{r},t)$ is the eastward component of the velocity.

3.09.4.1.1 Secular trend, tidal dissipation, and glacial isostatic adjustment

Tidal dissipation causes the Earth's angular velocity and hence rotational angular momentum to decrease. Since the angular momentum of the Earth–Moon system is conserved, the orbital angular momentum of the Moon must increase to balance the decrease in the Earth's rotational angular momentum. The increase in the orbital angular momentum of the Moon is accomplished by an increase in the radius of the Moon's orbit and a decrease in the Moon's orbital angular velocity. But early observations of the Moon's position showed that it was apparently accelerating, not decelerating, in its orbit. This apparent acceleration of the Moon was a result of assuming that the Earth is rotating with a constant, rather than decreasing, angular velocity when predicting the Moon's position. If the Earth's angular velocity is actually decreasing but is assumed to be constant when predicting the position of the Moon, then the observed position of the Moon will appear to be ahead of its predicted position. That is, the Moon will appear to be accelerating in its orbit. That the Moon is apparently accelerating in its orbit was first noted by Halley (1695). But it was not until 1939 that Spencer Jones (1939) was able to conclusively demonstrate that the angular velocity of the Earth is actually decreasing and that the apparent acceleration of the Moon in its orbit was an artifact of assuming that the angular velocity of the Earth was constant.

Halley (1695) also seemed to have been the first to appreciate the importance of ancient and medieval records of lunar and solar eclipses for determining the apparent acceleration of the Moon and the corresponding decrease in the angular velocity of the Earth over the past few thousand years. The change in the Earth's rate of rotation can be deduced from the discrepancy between when and where eclipses should have been observed if the angular velocity of the Earth were constant to when and where they were actually observed as recorded in Babylonian clay tablets and Chinese, European, and Arabic books and manuscripts (Stephenson, 1997, 1998; Stephenson and Morrison, 2005).

When using eclipse observations to deduce the secular change in the length of day over the past few thousand years, the positions of the Sun and Moon must be accurately known. Of primary importance in this regard is the value for the tidal acceleration $\dot{n}$ of the Moon since it controls the long-term behavior of the Moon's motion. The tidal acceleration of the Moon can be determined from observations of the timings of transits of Mercury (e.g., Morrison and Ward, 1975; Spencer Jones, 1939) and from satellite and lunar laser ranging measurements. Tidal forces distort the figure of the Earth and hence its gravitational field, which in turn perturbs the orbits of artificial satellites. SLR measurements can detect these tidal perturbations in the satellites' orbits and can therefore be used to construct tide models and hence determine the tidal acceleration of the Moon. Using this approach, Christodoulidis et al. (1988) reported a value of -25.27 ± 0.61 arcseconds per century^2 for the tidal acceleration $\dot{n}$ of the Moon due to dissipation by solid Earth and ocean tides. Other SLR-derived values for $\dot{n}$ have been reported by Cheng et al. (1990, 1992), Marsh et al. (1990, 1991), Dickman (1994), Lerch et al. (1994), and Ray (1994). Models of the solid Earth and ocean tides have been used to compute the concomitant secular increase in LOD by Ray et al. (1999), Wu et al. (2003), and Mathews and Lambert (2009).

Like the orbits of artificial satellites, the orbit of the Moon is also perturbed by tidal forces. Since LLR measurements can detect tidal perturbations in the Moon's orbit, they can be used

to determine the tidal acceleration of the Moon. In addition to being sensitive to orbital perturbations caused by tides on the Earth, LLR measurements, unlike SLR measurements, are also sensitive to orbital perturbations caused by tides on the Moon. Using LLR measurements, Williams et al. (2001) reported a value of -25.73 ± 0.5 arcseconds per century^2 for the tidal acceleration of the Moon, which by Kepler's law corresponds to an increase of 3.79 ± 0.07 cm year^{-1} in the semimajor axis of the Moon's orbit and which includes a contribution of +0.29 arcseconds per century^2 from dissipation within the Moon itself. Only about half of the discrepancy between the SLR- and LLR-derived values for $\dot{n}$ due to dissipation by just tides on the Earth is currently understood (Williams et al., 2001). Other LLR-derived values for $\dot{n}$ have been reported by Newhall et al. (1988), Dickey et al. (1994a), and Chapront et al. (2002).

By a priori adopting a value for the tidal acceleration $\dot{n}$ of the Moon, lunar and solar eclipse observations can be used to determine the secular increase in the length of day over the past few thousand years. The most recent rereductions of lunar and solar eclipse observations for LOD changes are those of Stephenson and Morrison (1995) and Morrison and Stephenson (2001). Besides using eclipse observations spanning from 700 BC to AD 1600, they also used lunar occultation observations spanning 1600–1955.5 and optical astrometric and space-geodetic measurements spanning 1955.5–90. Adopting a value of -26.0 arcseconds per century^2 for $\dot{n}$, Morrison and Stephenson (2001) found that the length of day has increased at a rate of $+1.80 \pm 0.1$ ms cy^{-1} on average during the past 2700 years (see **Figure 3**). In addition to a secular trend, Stephenson and Morrison (1995) and Morrison and Stephenson (2001) also found evidence for a fluctuation in the length of day that has a peak-to-peak amplitude of about 8 ms and a period of about 1500 years (**Figure 3**), although Dalmau (1997) had questioned its existence.

By the conservation of angular momentum, a tidal acceleration of the Moon of -26.0 arcseconds per century^2 should be accompanied by a +2.3 ms cy^{-1} increase in the length of day (Stephenson and Morrison, 1995). Since the observed increase in the length of day is only +1.8 ms cy^{-1} (Morrison and Stephenson, 2001), some other mechanism or combination of mechanisms must be acting to change the length of day by -0.5 ms cy^{-1}. By eqns [43] and [58], changes both in the axial component of relative angular momentum and in the polar moment of inertia of the Earth can cause LOD to change. A secular trend in the general circulation of fluids like the atmosphere and oceans, and hence in atmospheric and oceanic angular momentum, is unlikely to be sustained over the course of a few thousand years. In fact, using results from a 100-year run of the Hadley Centre general circulation model of the atmosphere, de Viron et al. (2004b) found that the modeled secular trend in atmospheric angular momentum (AAM) during 1870–1997 causes a secular trend in LOD of only +0.08 ms cy^{-1}.

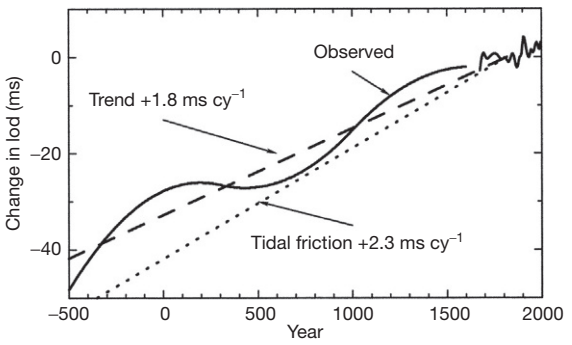


Figure 3 Secular change in the length of day during the past 2500 years estimated from lunar and solar eclipse, lunar occultation, optical astrometric, and space-geodetic observations. The difference between the observed secular trend and that caused by tidal friction is due to the effects of glacial isostatic adjustment and other processes such as ice sheet mass change and the accompanying nonsteric change in sea level. Reprinted from Journal of Geodynamics, 32, Morrison LV and Stephenson FR, Historical eclipses and the variability of the Earth's rotation. 247–265. Copyright (2001), with permission from Elsevier.

One of the most important mechanisms acting to cause a secular trend in LOD on timescales of a few thousand years is glacial isostatic adjustment (GIA). The isostatic adjustment of the solid Earth in response to the decreasing load on it following the last deglaciation causes the figure of the Earth to change and hence LOD to change. Since the solid Earth is rebounding in the regions at high latitude where the ice load was formerly located, the figure of the Earth is becoming less oblate, the Earth's rotation is accelerating, and LOD is decreasing. Models of GIA show that its effect on LOD is very sensitive to the assumed value of lower mantle viscosity (e.g., Johnston and Lambeck, 1999; Mitrovica and Milne, 1998; Mitrovica et al., 2006; Peltier and Drummond, 2010; Peltier and Jiang, 1996; Peltier and Luthcke, 2009; Peltier and Wu, 1983; Sabadini and Peltier, 1981; Sabadini and Vermeersen, 2004; Sabadini et al., 1982; Tamisea et al., 2002; Vermeersen et al., 1997; Wu and Peltier, 1984; Yuen et al., 1982). But by deriving a model for the radial viscosity profile of the Earth that fits both postglacial decay times and free-air gravity anomalies associated with mantle convection, Mitrovica and Forte (1997) found that GIA should cause a secular trend in LOD amounting to -0.5 ms cy^{-1}, a value in remarkable agreement with that needed to explain the difference between the observed secular trend in LOD and that caused by tidal dissipation.

However, GIA is not the only mechanism that will cause a secular trend in LOD. The present-day change in glacier and ice sheet mass and the accompanying change in nonsteric sea level will also cause a secular trend in LOD (e.g., James and Ivins, 1995, 1997; Mitrovica and Peltier, 1993; Mitrovica et al., 2006; Nakada and Okuno, 2003; Peltier, 1988; Tosi et al., 2005; Trupin, 1993; Trupin et al., 1992). But the effect of this mechanism on LOD is very sensitive to the still unknown present-day mass change of glaciers and ice sheets, particularly of the Antarctic ice sheet. By adopting various scenarios for the mass change of Antarctica, models predict that its mass change alone should cause a secular trend in LOD ranging anywhere from -0.72 ms cy^{-1} to +0.31 ms cy^{-1} (James and Ivins, 1997).

Other mechanisms that may cause a linear trend in LOD over the last or next century include continental drift (Dickman, 1979), tectonic processes taking place under non-isostatic conditions (Sabadini and Vermeersen, 2004; Vermeersen and Vlaar, 1993; Vermeersen et al., 1994), plate subduction (Alfonsi and Spada, 1998; Greff-Lefftz, 2011; Ricard et al., 1993; Spada et al., 1992), mantle convection (Greff-Lefftz, 2011), upwelling mantle plumes (Greff-Lefftz,

2011), deformation of the mantle caused by pressure variations acting at the core–mantle boundary that are associated with the motion of the fluid core (Dumberry and Bloxham, 2004; Fang et al., 1996; Greff-Lefftz et al., 2004), earthquakes (Chao and Gross, 1987; Gross and Chao, 2006), and climate change (Abarca del Rio, 1999; de Viron et al., 2002; Huang et al., 2001; Landerer et al., 2007; Ponte et al., 2002; Räisänen, 2003; Rosen and Gutowski Jr., 1992; Rosen and Salstein, 2000; Winkelnkemper et al., 2009).

The fluctuation in LOD of 1500-year period found by Stephenson and Morrison (1995) and Morrison and Stephenson (2001) is currently of unknown origin. However, given its large amplitude, which is too large to be caused by atmospheric and oceanic processes but which is comparable in size to the amplitude of the decadal variations, it is probably caused (Dumberry and Bloxham, 2006) by the same core–mantle interactions, such as gravitational coupling (Rubincam, 2003), that are known to cause decadal variations in LOD.

3.09.4.1.2 Decadal variations and core–mantle interactions

While lunar and solar eclipse observations are valuable for studying the secular trend in LOD, they are too sparse and inaccurate to reveal decadal variations. Instead, lunar occultation observations, which are available since the early 1600s (Martin, 1969; Morrison et al., 1981), are used to study decadal variations in LOD. **Figure 4** compares three different LOD series derived from lunar occultation observations. Gross (2001) (black curve) derived an LOD series spanning 1832.5–1997.5 at yearly intervals by analyzing UT1 measurements obtained from lunar occultation observations by Jordi et al. (1994), from the Hipparcos optical astrometric series of Vondrák (1991, 1999) and Vondrák et al. (1992, 1995, 1997, 1998), and from the COMB97 combined optical astrometric and space-geodetic series of Gross (2000a). McCarthy and Babcock (1986) (red curve) derived an LOD series spanning 1657.0–1984.5 at half-yearly intervals by analyzing UT1 measurements obtained from lunar occultation observations by Martin (1969) and Morrison (1979), from the optical astrometric series of McCarthy (1976), and from a series obtained from the BIH. Stephenson and Morrison (1984) (green curve) obtained an LOD series spanning 1630–1980 at 5-year intervals before 1780 and at yearly intervals afterward by analyzing the lunar occultation and solar eclipse observations cataloged by Morrison (1978) and Morrison et al. (1981), combining them with an LOD series derived from optical astrometric measurements of UT1 obtained from the BIH. As can be seen from **Figure 4**, during their common time span, these three different LOD series are consistent with each other to within the 1σ standard error of the Gross (2001) series. Other LOD series that have been derived from lunar occultation observations are those of Morrison (1979), Jordi et al. (1994), and Liao and Greiner-Mai (1999).

The peak-to-peak variation in LOD seen in **Figure 4** is about 7 ms, a variation far too large to be caused by atmospheric and oceanic processes (Gross et al., 2005). That atmospheric processes cannot cause such large decadal-scale LOD variations can be easily demonstrated by considering the change in LOD that would be caused if the motion of the atmosphere were to stop entirely. Because of the pole-to-equator temperature gradient, the atmosphere superrotates with respect to the solid Earth at an average rate of about 7 m s^{-1}. If the superrotation of the atmosphere were to stop so that the atmosphere just passively corotates with the solid Earth, then by conservation of angular momentum, the length of day would decrease by about 3 ms (Hide et al., 1980). Since stopping the atmospheric motion entirely does not cause an LOD change as large as that observed in **Figure 4**, some other mechanism or combination of mechanisms must be acting to cause the large decadal-scale LOD variations that are observed.

The most important mechanism acting to cause decadal variations in LOD is core–mantle coupling. While it has been recognized for quite some time that the core is the only viable

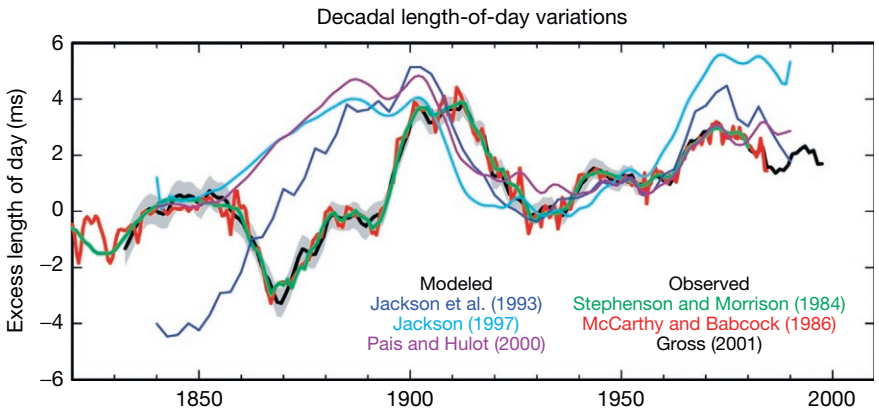


Figure 4 Plots of observed and modeled LOD variations on decadal timescales. The observed LOD series are those of Gross (2001) (black curve with gray shading representing $\pm 1\sigma$ standard error), McCarthy and Babcock (1986) (red curve), and Stephenson and Morrison (1984) (green curve). Note that after about 1955, the uncertainties of the Gross (2001) LOD values are less than the width of the black line. The modeled core angular momentum series are those of Jackson et al. (1993) (dark blue curve), the *uvm-s* model of Jackson (1997) (light blue curve), and the PH-inversion model of Pais and Hulot (2000) (purple curve). A secular trend of +1.8 ms cy^{-1} has been added to the modeled core angular momentum series to match the observed secular trend (see **Section 3.09.4.1.1**). An arbitrary bias has also been added to the modeled series in order to facilitate their comparison with the observed series. Reprinted from Physics of the Earth and Planetary Interiors, 123, Gross RS, A combined length-of-day series spanning 1832–1997: LUNAR97, 65–76. Copyright (2001), with permission from Elsevier.

source of the large decadal LOD variations that are observed (e.g., Lambeck, 1980; Munk and MacDonald, 1960), it was not until 1988 that Jault et al. (1988) were able to model the core angular momentum (CAM) and show that it causes decadal LOD variations that agree reasonably well with those observed.

The flow of the fluid at the top of the core can be inferred from surface observations of the secular variation of the magnetic field by assuming that (1) the mantle is a perfect insulator so that the magnetic field can be expressed as the gradient of a potential, which facilitates the downward continuation of the magnetic field to the top of the core; (2) the core is a perfect conductor so that the magnetic field is 'frozen' into the core fluid and is thus advected by the horizontal flow at the top of the core; (3) the flow at the top of the core is tangentially geostrophic so that it is governed by the balance between the horizontal components of just the pressure gradient and the Coriolis force at the top of the core (Bloxham and Jackson, 1991; Whaler and Davis, 1997); and (4) the flow is large scale. The last two assumptions are required in order to reduce the inherent nonuniqueness of core surface flow determinations. Other assumptions about the core surface flow fields that have been made (Finlay et al., 2010; Holme, 2007) are that the flow is purely toroidal so that it has no radial component (Bloxham, 1990; Whaler, 1980, 1986), that the flow is steady in time (Gubbins, 1982; Voorhies, 1986; Voorhies and Backus, 1985), that the flow is steady within a drifting reference frame (Davis and Whaler, 1996; Holme and Whaler, 2001), that the flow includes a helical component (Amit and Olson, 2004, 2006), that the flow is tangentially magnetostrophic (Asari and Lesur, 2011), or that the flow is quasi-geostrophic (Canet et al., 2009; Gillet et al., 2009; Pais and Jault, 2008).

While surface magnetic field observations can be used to infer the flow at the top of the core, the flow everywhere within the core must be known in order to compute the angular momentum of the core and hence the effect of core motion on the length of day. Jault et al. (1988) realized from dynamical considerations that for geostrophic flows on decadal timescales, the axisymmetric, or zonal, component of the core flow can be described by relative motion of nested cylinders that are coaxial with the rotation axis. This model for the motion of the core at depth allows the axial angular momentum of the core, and hence the effect of core motion on LOD, to be computed from the flow fields at the top of the core that are inferred from surface magnetic field observations (Amit and Olson, 2006; Asari et al., 2009; Gillet et al., 2009; Hide et al., 2000; Holme and Olsen, 2006; Holme and Whaler, 2001; Jackson, 1997; Jackson et al., 1993; Jault et al., 1988; Olsen and Mandea, 2008; Pais and Hulot, 2000; Pais and Jault, 2008; Pais et al., 2004; Wardinski and Lesur, 2012; Wardinski et al., 2008; Whaler and Holme, 2011). **Figure 4** compares the observed decadal LOD variations with the modeled results obtained by Jackson et al. (1993), Jackson (1997), and Pais and Hulot (2000). As can be seen, while the agreement is not perfect, CAM models produce decadal LOD variations that have about the same amplitude and phase as those observed.

In the coaxial nested cylinder model of the zonal core flow, the rotation of adjacent cylinders is coupled because of the magnetic field lines that thread through them. If for some reason the rotation of the cylinders is disturbed, then the magnetic field provides a restoring force that will cause the cylinders to oscillate. As first proposed by Braginsky (1970), it is the exchange with the mantle of the angular momentum associated with these torsional oscillations of the fluid core that causes the length of day to change (Buffett, 1998; Buffett and Mound, 2005; Buffett et al., 2009; Dickey and de Viron, 2009; Hide et al., 2000; Mound and Buffett, 2005, 2007; Pais and Hulot, 2000; Zatman and Bloxham, 1997, 1998, 1999).

While the exchange of CAM with the solid Earth can clearly cause decadal LOD variations of approximately the right amplitude and phase, the mechanism or mechanisms by which the angular momentum is exchanged between the core and solid Earth are less certain. Possible core–mantle coupling mechanisms are viscous torques, topographic torques, electromagnetic torques, and gravitational torques (Bloxham, 1998; Buffett, 2007; Jault, 2003; Kuang and Chao, 2003; Ponsar et al., 2003; Roberts and Aurnou, 2012). Viscous coupling is caused by the drag of the core flow on the core–mantle boundary, with the strength of the coupling depending on the viscosity of the core fluid. Given the current estimates of core viscosity, it is generally agreed that viscous torques are too weak to be effective in coupling the core to the mantle (Rochester, 1984).

If the core–mantle boundary is not smooth but exhibits undulations or 'bumps,' then the flow of the core fluid can exert a torque on the mantle due to the fluid pressure acting on the boundary topography (Asari et al., 2006; Buffett, 1998; Hide, 1969, 1977, 1989, 1993, 1995a; Hide et al., 1993; Hinderer et al., 1990; Jault and Le Mouël, 1989, 1990, 1991; Jault et al., 1996; Kuang and Bloxham, 1993; Kuang and Chao, 2001; Mound and Buffett, 2005). The strength of this topographic coupling, a mechanism first suggested by Hide (1969), depends on the amplitude of the topography at the core–mantle boundary. Because of uncertainties in the size of this topography and a controversy about how the topographic torque should be computed (Bloxham and Kuang, 1995; Hide, 1995b, 1998; Jault and Le Mouël, 1999; Kuang and Bloxham, 1997), there is as yet no consensus on the importance of topographic coupling as a mechanism for exchanging angular momentum between the core and mantle.

Electromagnetic torques arise from the interaction between the magnetic field within the core and the flow of electric currents in the weakly conducting mantle that are induced both by time variations of the magnetic field and by diffusion of electric currents from the core to the mantle (Bullard et al., 1950; Dumberry and Mound, 2008; Holme, 1998a,b, 2000; Jault and Le Mouël, 1991; Love and Bloxham, 1994; Mound and Buffett, 2005; Nakada, 2006, 2009a, 2011; Roberts, 1972; Rochester, 1960, 1962; Roden, 1963; Stewart et al., 1995; Stix and Roberts, 1984; Wicht and Jault, 1999, 2000). The strength of this electromagnetic torque, a mechanism first suggested by Bullard et al. (1950), depends both on the conductivity of the mantle and on the strength of the magnetic field crossing the core–mantle boundary. If the conductivity of the mantle, or of a narrow layer at the base of the mantle, is sufficiently large, then electromagnetic torques can produce decadal LOD variations as large as those observed. But because of uncertainties in the conductivity at the base of the mantle, the importance of electromagnetic coupling, like that of topographic coupling, as a mechanism for exchanging angular momentum between the core and mantle remains unclear.

Gravitational attraction between density heterogeneities in the fluid core and mantle can exert a torque on the mantle, leading to changes in the length of day (Buffett, 1996a; Jault and Le Mouël, 1989; Xu et al., 2000). The strength of the gravitational torque depends upon the size of the mass anomalies in the core and mantle, which are poorly known. As a result, there have been few quantitative estimates of the magnitude of the gravitational torque. However, Buffett (1996a,b) suggested that the inner core may be gravitationally locked to the mantle. If so, then any rotational disturbance of the inner core, possibly caused by electromagnetic torques acting on the inner core, will be transmitted to the mantle, causing LOD variations. While Buffett (1998) and Mound and Buffett (2005) considered this last mechanism to be the most viable mechanism for exchanging angular momentum between the core and mantle, Zatman (2003) found that it is inconsistent with a model of inner core rotation rate determined from the core flow within the tangent cylinder. In any case, the amplitude of the observed decadal LOD variations has been used to place limits on the strength of the gravitational coupling between the inner core and the mantle (Aubert and Dumberry, 2011; Buffett and Creager, 1999; Dumberry, 2007; Dumberry and Mound, 2010).

Geomagnetic jerks are sudden changes in the otherwise smoothly changing secular variation of the geomagnetic field. They manifest themselves either as changes in the slope of the secular variation or, equivalently, as steplike changes in the secular acceleration (Mandea et al., 2010). The occurrence of geomagnetic jerks appears to be correlated with sudden changes in the time rate of change of LOD (e.g., Holme and de Viron, 2005; Mandea et al., 2010; Olsen and Mandea, 2007), with changes in the phase of the Chandler wobble (Bellanger et al., 2001, 2002a; Gibert and Le Mouël, 2008; Gibert et al., 1998), and with changes in the phase of the free core nutation (Shirai et al., 2005). While the origin of geomagnetic jerks is currently unknown, these correlations suggest that their cause is related to the same processes in the core that cause the angular momentum of the core, and hence the rotation of the Earth, to change such as torsional oscillations (Bloxham et al., 2002).

3.09.4.1.3 Tidal variations and solid Earth, oceanic, and atmospheric tides

Tidal forces due to the gravitational attraction of the Sun, Moon, and planets deform the solid and fluid parts of the Earth, causing the Earth's inertia tensor to change and hence the Earth's rotation to change. Jeffreys (1928) was the first to predict that the periodic displacement of the solid and fluid masses of the Earth associated with the tides should cause periodic changes in the Earth's rate of rotation at the tidal frequencies. While Markowitz (1955, 1959) reported observing such periodic variations at the fortnightly and monthly tidal frequencies from observations taken at two photographic zenith tubes, these observations were later thrown into doubt by the error analysis of Fliegel and Hawkins (1967). By combining observations from about 55 instruments, Guinot (1970) was able to detect variations in the Earth's rate of rotation at the fortnightly and monthly tidal frequencies that had amplitudes significantly greater than the level of observation noise. Today, the high accuracy of the space-geodetic measurement systems allows long-period tidal effects on UT1 and LOD to be unambiguously observed (e.g., Bellanger et al., 2002b; Benjamin et al., 2006; Chao et al., 1995a; Dickman and Nam, 1995; Gross, 2009a; Hefty and Capitaine, 1990; McCarthy and Luzum, 1993; Nam and Dickman, 1990; Ray and Egbert, 2012; Robertson et al., 1994; Schastok et al., 1994; Schuh and Schmitz-Hübsch, 2000).

In a seminal paper, Yoder et al. (1981) derived a model for the long-period tidal variations in the rotation of the Earth, assuming that the crust and mantle of the Earth are elastic, that the core is decoupled from the mantle, and that the ocean tides are in equilibrium. **Table 5** gives their results for the long-period tidal variations in UT1 and LOD, where the LOD results have been derived here from their UT1 results by using eqn [64]. On intraseasonal timescales, the largest effects are found to be at the fortnightly and monthly tidal periods. The large tidal variations at annual and semiannual periods are obscured in Earth rotation observations by meteorologic effects (see **Section 3.09.4.1.4**), while those having periods of 9.3 and 18.6 years are obscured by the decadal variations in the Earth's rotation (see **Section 3.09.4.1.2**).

The effects of mantle anelasticity on the tidal variations in the Earth's rate of rotation have been discussed by Merriam (1984, 1985), Wahr and Bergen (1986), Defraigne and Smits (1999), Benjamin et al. (2006), and Ray and Egbert (2012). Dissipation associated with mantle anelasticity causes the deformational and hence rotational response of the Earth to lag behind the forcing tidal potential. As a result, not only does mantle anelasticity modify the in-phase rotational response of the Earth to the tidal potential, but also out-of-phase terms are introduced. Anelastic effects are found to modify the elastic rotational response of the Earth by a few percent.

Defraigne and Smits (1999) also considered the effects of nonhydrostatic structure within the Earth on the tidal variations in the Earth's rotation. A nonhydrostatic initial state of the Earth was determined by computing the buoyancy-driven flow in the mantle due to the seismically observed mass anomalies there while accounting for the associated flow-induced boundary deformation, potential readjustment, and mass readjustment in the outer and inner cores. Their results indicate that nonhydrostatic structure within the Earth modifies the rotational response of the Earth to the zonal tide generating potential by less than 0.1%.

Dynamic effects of long-period ocean tides on the Earth's rotation using ocean tide models based upon Laplace's tidal equations have been computed by Brosche et al. (1989), Seiler (1990, 1991), Wünsch and Busshoff (1992), Dickman (1993), Gross (1993), Seiler and Wünsch (1995), Benjamin et al. (2006), Weis (2006), and Dickman and Gross (2010). But the accuracy of ocean tide models greatly improved when TOPEX/Poseidon (T/P) sea surface height measurements became available (Gross, 2009a). Dynamic effects of long-period ocean tides on the Earth's rotation using tide models based upon T/P sea surface height measurements have been computed by Kantha et al. (1998), Desai and Wahr (1999), and Ray and Egbert (2012). **Table 6** gives the results obtained by Kantha et al. (1998) from their ocean tide model that assimilates T/P-derived tides. As expected, dynamic tide effects are seen to be larger at the fortnightly tidal frequency than they are at the monthly frequency, with the amplitude of the

Table 5 Modeled variations in UT1 and LOD caused by elastic solid body and equilibrium ocean tides

Argument					Period	ΔUT1	ΔΛ(t)	Argument					Period	ΔUT1	ΔΛ(t)
l	l′	F	D	Ω	(days)	sin (μs)	cos (μs)	l	l′	F	D	Ω	(days)	sin (μs)	cos (μs)
1	0	2	2	2	5.64	−2.35	2.62	1	0	0	0	−1	27.44	53.39	−12.22
2	0	2	0	1	6.85	−4.04	3.71	1	0	0	0	0	27.56	−826.07	188.37
2	0	2	0	2	6.86	−9.87	9.04	1	0	0	0	1	27.67	54.43	−12.36
0	0	2	2	1	7.09	−5.08	4.50	0	0	0	1	0	29.53	4.70	−1.00
0	0	2	2	2	7.10	−12.31	10.90	1	−1	0	0	0	29.80	−5.55	1.17
1	0	2	0	0	9.11	−3.85	2.66	−1	0	0	2	−1	31.66	11.75	−2.33
1	0	2	0	1	9.12	−41.08	28.30	−1	0	0	2	0	31.81	−182.36	36.02
1	0	2	0	2	9.13	−99.26	68.29	−1	0	0	2	1	31.96	13.16	−2.59
3	0	0	0	0	9.18	−1.79	1.22	1	0	−2	2	−1	32.61	1.79	−0.34
−1	0	2	2	1	9.54	−8.18	5.38	−1	−1	0	2	0	34.85	−8.55	1.54
−1	0	2	2	2	9.56	−19.74	12.98	0	2	2	−2	2	91.31	−5.73	0.39
1	0	0	2	0	9.61	−7.61	4.98	0	1	2	−2	1	119.61	3.29	−0.17
2	0	2	−2	2	12.81	2.16	−1.06	0	1	2	−2	2	121.75	−188.47	9.73
0	1	2	0	2	13.17	2.54	−1.21	0	0	2	−2	0	173.31	25.10	−0.91
0	0	2	0	0	13.61	−29.89	13.80	0	0	2	−2	1	177.84	117.03	−4.13
0	0	2	0	1	13.63	−320.82	147.86	0	0	2	−2	2	182.62	−4824.74	166.00
0	0	2	0	2	13.66	−775.69	356.77	0	2	0	0	0	182.63	−19.36	0.67
2	0	0	0	−1	13.75	2.16	−0.99	2	0	0	−2	−1	199.84	4.89	−0.15
2	0	0	0	0	13.78	−33.84	15.43	2	0	0	−2	0	205.89	−54.71	1.67
2	0	0	0	1	13.81	1.79	−0.81	2	0	0	−2	1	212.32	3.67	−0.11
0	−1	2	0	2	14.19	−2.44	1.08	0	−1	2	−2	1	346.60	−4.51	0.08
0	0	0	2	−1	14.73	4.70	−2.00	0	1	0	0	−1	346.64	9.21	−0.17
0	0	0	2	0	14.77	−73.41	31.24	0	−1	2	−2	2	365.22	82.81	−1.42
0	0	0	2	1	14.80	−5.26	2.24	0	1	0	0	0	365.26	−1535.87	26.42
0	−1	0	2	0	15.39	−5.08	2.07	0	1	0	0	1	386.00	−13.82	0.22
1	0	2	−2	1	23.86	4.98	−1.31	1	0	0	−1	0	411.78	3.48	−0.05
1	0	2	−2	2	23.94	10.06	−2.64	2	0	−2	0	0	−1095.17	−13.72	−0.08
1	1	0	0	0	25.62	3.95	−0.97	−2	0	2	0	1	1305.47	42.11	−0.20
−1	0	2	0	0	26.88	4.70	−1.10	−1	1	0	1	0	3232.85	−4.04	0.01
−1	0	2	0	1	26.98	17.67	−4.11	0	0	0	0	2	−3399.18	789.98	1.46
−1	0	2	0	2	27.09	43.52	−10.09	0	0	0	0	1	−6798.38	−161726.81	−149.47

The tabulated coefficients are derived using a constant value of $k/C = 0.94$ as recommended by Yoder et al. (1981). Terms with UT1 amplitudes less than 2 μs are not tabulated. *l*, *l′*, *F*, *D*, and Ω are the Delaunay arguments, expressions for which are given in Simon et al. (1994). The period, given in solar days, is the approximate period of the term as viewed in the terrestrial reference frame.

Source: Yoder CF, Williams JG, and Parke ME (1981) Tidal variations of Earth rotation. *Journal of Geophysical Research* 86: 881–891.

Table 6 Modeled variations in UT1 and LOD caused by long-period dynamic ocean tides

Tide	Argument					Period (days)	ΔUT1 mass (μs)		ΔUT1 motion (μs)		ΔΛ(t) mass (μs)		ΔΛ(t) motion (μs)	
	l	l′	F	D	Ω		sin	cos	sin	cos	cos	sin	cos	sin
M_f	0	0	2	0	2	13.66	−102.8	33.0	−1.4	15.4	47.28	15.18	0.64	7.08
M_m	1	0	0	0	0	27.56	−119.1	8.8	5.7	10.7	27.15	2.01	−1.30	2.44

The tabulated coefficients are from the assimilated (A) model of Kantha et al. (1998). *l*, *l′*, *F*, *D*, and Ω are the Delaunay arguments, expressions for which are given in Simon et al. (1994). The period, given in solar days, is the approximate period of the term as viewed in the terrestrial reference frame.

Source: Kantha LH, Stewart JS, and Desai SD (1998) Long-period lunar fortnightly and monthly ocean tides. *Journal of Geophysical Research* 103(C6): 12639–12647.

out-of-phase mass and motion terms (the cosine coefficients for UT1 and the sine coefficients for LOD) each being larger for the fortnightly tide than they are for the monthly tide.

Ocean tides in the diurnal and semidiurnal tidal bands also affect the Earth's rate of rotation. While Yoder et al. (1981) were the first to predict these effects using theoretical ocean tide models based upon Laplace's tidal equations, they were not actually observed until Dong and Herring (1990) detected them in VLBI measurements. Subdaily variations in UT1 and LOD at the diurnal and semidiurnal tidal frequencies have now been unambiguously observed in measurements taken by VLBI (Artz et al., 2010, 2011; Böhm et al., 2012; Brosche et al., 1991; Gipson, 1996; Haas and Wünsch, 2006; Herring, 1993; Herring and Dong, 1991, 1994; Sovers et al., 1993; Wünsch and Busshoff, 1992), SLR (Watkins and Eanes, 1994), GPS (Freedman et al., 1994b; Grejner-Brzezinska and Goad, 1996;

Hefty et al., 2000; Lichten et al., 1992; Malla et al., 1993; Rothacher et al., 2001; Steigenberger et al., 2006), and in combinations of GPS and VLBI (Artz et al., 2012; Thaller et al., 2005, 2006, 2007).

Following Yoder et al. (1981), predictions of subdaily tidal variations in the Earth's rate of rotation using theoretical ocean tide models have been made by Brosche (1982), Baader et al. (1983), Brosche et al. (1989), Seiler (1990, 1991), Wünsch and Busshoff (1992), Dickman (1993), Gross (1993), Seiler and Wünsch (1995), and Weis (2006). But as with long-period ocean tide models, the accuracy of diurnal and semidiurnal tide models greatly improved when T/P sea surface height measurements became available. Dynamic effects of diurnal and semidiurnal ocean tides on the Earth's rotation computed from tide models incorporating T/P sea surface height measurements have been given by Ray et al. (1994), Chao et al. (1995b, 1996a), and Chao and Ray (1997). Extensive tables of coefficients for the effect of diurnal and semidiurnal ocean tides on UT1 and LOD are given by Petit and Luzum (2010) (**Chapter 1.08**).

Comparisons of observations with models show the dominant role that ocean tides play in causing subdaily UT1 and LOD variations (**Figure 5(a)**), with as much as 90% of the observed UT1 variance being explained by diurnal and semidiurnal ocean tides (Chao et al., 1996a; Thaller et al., 2006). Apart from errors in observations and models, the small difference that remains (e.g., Schindelegger et al., 2011; Schuh and Schmitz-Hübsch, 2000) may be due to nontidal atmospheric and oceanic effects.

The diurnally varying solar heating of the atmosphere excites diurnal and semidiurnal tidal waves in the atmosphere that travel westward with the Sun (Chapman and Lindzen, 1970; Dai and Wang, 1999; Haurwitz and Cowley, 1973; Volland, 1988, 1997). These radiational tides are much larger than the gravitational tides in the atmosphere, with the amplitude of the surface pressure variations due to the radiational tides being about 20 times larger than the amplitude due to the gravitational tides. While gravitational tides in the atmosphere have no discernible effect on the Earth's rotation, the radiational tides do have an effect (Brzezinski et al., 2002a; de Viron et al., 2005; Zharov and Gambis, 1996). Using the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis wind and pressure fields, Brzezinski et al. (2002a) predicted that the UT1 variations caused by the diurnal S_1 radiational tide should have an amplitude of about 0.5 μs and that the LOD variations should have an amplitude of about 3.3 μs. Since the NCEP/NCAR reanalysis wind and pressure fields are given every 6 h, the semidiurnal tidal frequency band is incompletely sampled, making it difficult to estimate the effect on UT1 and LOD of the semidiurnal S_2 radiational tide. In addition, since the oceans will respond dynamically to the tidal variations in the atmospheric wind and pressure fields, the oceans will also contribute to the excitation of UT1 and LOD by the radiational tides. In fact, the effect of radiational tides on UT1 and LOD is typically included in the

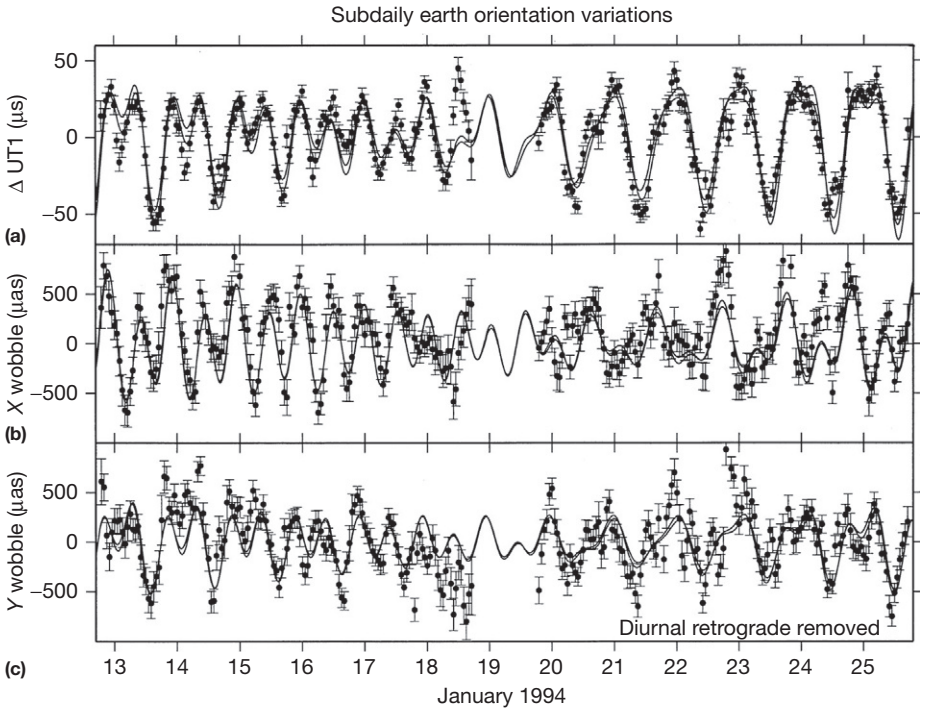


Figure 5 Plots of observed and modeled (a) UT1, (b) *x*-component of polar motion, and (c) *y*-component of polar motion during the Cont94 measurement campaign of 12–26 January 1994. The dots with 1σ error bars are the hourly VLBI observations. Polar motion variations in the retrograde nearly diurnal frequency band have been removed from the observed series and are not included in the modeled series. The solid lines are the predicted effects from the diurnal and semidiurnal T/P ocean tide models B and C of Chao et al. (1996a). Each tide model explains about 90% of the observed UT1 variance and about 60% of the observed polar motion variance. Reproduced from Chao BF, Ray RD, Gipson JM, Egbert GD, and Ma C (1996a) Diurnal/semidiurnal polar motion excited by oceanic tidal angular momentum. *Journal of Geophysical Research* 101(B9): 20151–20163.

tables of the effects of diurnal and semidiurnal ocean tides on the Earth's rate of rotation (see, e.g., Petit and Luzum, 2010, Table 8.3a and 8.3b).

3.09.4.1.4 Seasonal variations

Seasonal variations in LOD were first detected by Stoyko (1937). Numerous studies have since shown that the observed annual and semiannual variations in LOD are primarily caused by annual and semiannual changes in the angular momentum of the zonal winds (e.g., Aoyama and Naito, 2000; Eubanks et al., 1985; Eubanks, 1993; Hide and Dickey, 1991; Höpfner, 1998, 2000b; Lambeck, 1980, 1988; Lambeck and Hiopgood, 1981; Munk and MacDonald, 1960; Neef and Matthes, 2012; Rosen, 1993; Yu et al., 1999; Zhou et al., 2008). In fact, within the uncertainty of the earlier LOD measurements, seasonal variations in LOD could be accounted for solely by seasonal variations in the zonal winds (Dickey et al., 1993a; Naito and Kikuchi, 1990, 1991; Rosen, 1993; Rosen and Salstein, 1985, 1991). But as measurement accuracies improved, discrepancies between the observed and modeled variations became noticeable.

With the advent of models of the general circulation of the oceans, it became possible to evaluate the effect of the oceans on seasonal LOD variations (Brosche and Sündermann, 1985; Brosche et al., 1990, 1997; Dickey et al., 1993b; Frische and Sündermann, 1990; Gross, 2004; Gross et al., 2004; Johnson et al., 1999; Kouba and Vondrák, 2005; Marcus et al., 1998; Ponte and Stammer, 2000; Ponte et al., 2001, 2002; Segschneider and Sündermann, 1997; Yan et al., 2006). For example, **Table 7** gives the results obtained by Gross et al. (2004). They found that during 1992–2000, the amplitude of the observed annual LOD variation was 369.0 ± 6.4 μs with that caused by the effect of zonal winds integrated to a height of 10 hectoPascals (hPa) being 414.8 ± 5.0 μs, by atmospheric surface pressure variations being 37.4 ± 0.8 μs, by oceanic currents being 7.6 ± 0.3 μs, and by ocean-bottom pressure variations being 7.9 ± 0.3 μs. The sum of the effects of atmospheric winds to 10 hPa, surface pressure, oceanic currents, and bottom pressure had an annual amplitude of 363.0 ± 5.2 μs, the same as that observed to within measurement uncertainty, and a phase difference of only 4.5°. But the effects of the winds above 10 hPa have not yet been taken into account.

Although only 1% of the atmospheric mass is located in the region of the atmosphere above 10 hPa, the strength of the zonal winds there is great enough that they have a noticeable effect on seasonal LOD variations (Dickey et al., 1993a; Höpfner, 2001; Rosen, 1993; Rosen and Salstein, 1985, 1991). Gross et al. (2004) found that during 1992–2000, the annual LOD variation caused by winds above 10 hPa had an amplitude of 20.5 ± 0.3 μs, larger than the sum of the effects of oceanic currents and bottom pressure. Since the annual stratospheric winds are out of phase with the annual tropospheric winds, including the effect of the winds above 10 hPa brings the amplitude of the modeled annual LOD variations down to 343.5 ± 5.2 μs. Thus, when the effects of stratospheric winds are included, the annual LOD budget is not closed.

Table 7 also gives the results obtained by Gross et al. (2004) for the semiannual and terannual (3 cpy) LOD variations. Like the annual LOD budget, the semiannual budget is also not closed. But within measurement uncertainty, LOD variations at the terannual frequency are completely accounted for by the effects of the atmosphere and oceans.

Apart from errors in observations and models, the residual that remains after modeled atmospheric and oceanic effects

Table 7 Observed and modeled nontidal LOD variations at seasonal frequencies during 1992–2000

Excitation process	*Annual*		*Semiannual*		*Terannual*	
	Amplitude (μs)	*Phase (°)*	*Amplitude (μs)*	*Phase (°)*	*Amplitude (μs)*	*Phase (°)*
Observed	369.0 ± 6.4	31.6 ± 1.0	294.3 ± 6.3	-116.5 ± 1.2	52.4 ± 6.4	20.9 ± 7.0
Atmospheric						
Winds (ground to 10 hPa)	414.8 ± 5.0	34.7 ± 0.7	244.3 ± 5.0	-110.0 ± 1.2	54.1 ± 5.0	30.4 ± 5.3
Winds (10–0.3 hPa)	20.5 ± 0.3	-161.0 ± 0.9	29.4 ± 0.3	-122.7 ± 0.6	3.5 ± 0.3	-165.7 ± 5.1
All winds (ground to 0.3 hPa)	395.1 ± 5.0	35.5 ± 0.7	273.1 ± 5.0	-111.3 ± 1.0	50.7 ± 5.0	31.4 ± 5.6
Surface pressure (IB)	37.4 ± 0.8	-154.6 ± 1.3	9.2 ± 0.8	113.3 ± 5.1	2.6 ± 0.8	-48.8 ± 18.1
All winds and surface pressure	358.3 ± 5.1	36.5 ± 0.8	266.7 ± 5.1	-112.7 ± 1.1	51.2 ± 5.1	28.6 ± 5.7
Oceanic						
Currents	7.6 ± 0.3	-165.5 ± 2.3	0.9 ± 0.3	115.5 ± 18.7	1.6 ± 0.3	35.9 ± 11.1
Bottom pressure	7.9 ± 0.3	-148.4 ± 2.2	3.1 ± 0.3	176.7 ± 5.5	1.0 ± 0.3	-67.3 ± 16.4
Currents and bottom pressure	15.3 ± 0.5	-156.8 ± 2.0	3.6 ± 0.5	163.4 ± 8.6	1.7 ± 0.5	-0.6 ± 18.4
Atmospheric and oceanic						
All winds and currents	388.0 ± 5.0	35.9 ± 0.7	272.5 ± 5.0	-111.5 ± 1.1	52.3 ± 5.0	31.6 ± 5.5
Surface and bottom pressure	45.2 ± 0.9	-153.5 ± 1.2	10.9 ± 0.9	127.9 ± 4.9	3.6 ± 0.9	-54.1 ± 15.1
Total of all atmospheric and oceanic						
Without winds above 10 hPa	363.0 ± 5.2	36.1 ± 0.8	238.1 ± 5.2	-112.4 ± 1.3	56.1 ± 5.2	26.9 ± 5.3
With winds above 10 hPa	343.5 ± 5.2	37.1 ± 0.9	267.1 ± 5.2	-113.5 ± 1.1	52.7 ± 5.2	27.7 ± 5.7

IB, inverted barometer. The amplitude A and phase α of the observed and modeled seasonal LOD variations are defined by $\Delta\Lambda(t) = A\cos[\sigma(t - t_o) - \alpha]$ where σ is the annual, semiannual, or terannual frequency and the reference date t_o is 1 January 1990.

Source: Gross RS, Fukumori I, Menemenlis D, and Gegout P (2004) Atmospheric and oceanic excitation of length-of-day variations during 1980–2000. *Journal of Geophysical Research* 109: B01406, http://dx.doi.org/10.1029/2003JB002432.

have been removed from the observations may be caused by hydrologic processes (Chao and O'Conner, 1988; Chen, 2005; Chen et al., 2000a; Yan and Chao, 2012; Zhong et al., 2003). For example, Chen (2005) found that during 1993.0–2004.3, modeled hydrologic processes cause an annual change in LOD of amplitude 17.2 μs, which when added to oceanic effects is reduced to 12.5 μs. However, when the effects of balancing mass within the atmosphere, ocean, and hydrologic system are included, this is further reduced to 1.4 μs, which is not large enough to close the annual LOD budget. His results for the semiannual LOD variations are similar. However, Yan and Chao (2012) found that the annual budget can be nearly closed when global mass balance is enforced and when, in addition, the effect of atmospheric winds is inflated by about 7% (Chao and Yan, 2010).

Since meteorologic processes are the predominant cause of seasonal LOD variations and since these processes can change from year to year, there is no reason to expect that the seasonal LOD variations should be the same from year to year, either in amplitude or in phase. In fact, Feissel and Gavoret (1990) showed that the amplitude of the annual LOD oscillation was about twice as large as normal during the 1982–83 El Niño/Southern Oscillation (ENSO) event, with the amplitude of the semiannual LOD oscillation being about half as large as normal. Gross et al. (1996a, 2002) extended this study, showing that during 1962–2000, the amplitudes of the seasonal LOD and wind-driven AAM variations at both annual and semiannual frequencies have not been constant but have fluctuated by as much as 50%. They also showed that the changing amplitudes of the annual and semiannual LOD and AAM variations are significantly correlated with the Southern Oscillation Index (SOI), an index defined to be the normalized difference in surface pressure between Darwin and Tahiti. Since the SOI is also correlated with LOD and AAM variations occurring on interannual timescales (see **Section 3.09.4.1.5**), the significant correlation they observed between changes in the amplitude of the seasonal cycle and the SOI is evidence of a linkage between the seasonal cycle and interannual LOD and AAM variations, a linkage that can only occur through nonlinear interactions.

3.09.4.1.5 Interannual variations and ENSO

Like seasonal variations in LOD, variations on interannual timescales are also predominantly caused by changes in the angular momentum of the zonal winds (e.g., Eubanks, 1993; Hide and Dickey, 1991; Lambeck and Hopgood, 1981; Neef and Matthes, 2012; Rosen, 1993; Yu et al., 1999). The most prominent feature of the climate system on these timescales is the ENSO phenomenon. ENSO is a global-scale oscillation of the coupled atmosphere–ocean system characterized by fluctuations in atmospheric surface pressure and ocean temperatures in the tropical Pacific (e.g., Philander, 1990). During an ENSO event, in which the SOI decreases, the tropical easterlies collapse causing the AAM to increase. By the conservation of angular momentum, as the AAM increases, the solid Earth's angular momentum decreases and the length of day increases. Numerous studies (e.g., Abarca del Rio et al., 2000; Chao, 1984, 1988, 1989; Dickey et al., 1992a, 1993a, 1994b, 1999; Eubanks et al., 1986; Feissel and Gavoret, 1990; Gambis, 1992; Jordi et al., 1994; Rosen et al., 1984; Salstein and Rosen, 1986; Stefanick, 1982; Zheng et al., 2003; Zhou et al., 2001) have shown that observed LOD variations on interannual timescales, as well as interannual variations in the angular momentum of the zonal winds, are (negatively) correlated with the SOI, reflecting the impact on the length of day of changes in the zonal winds associated with ENSO. For example, Chao (1984) reported finding a correlation between the interannual LOD variations and the SOI during 1957–83 with a maximum negative value for the correlation coefficient of −0.56 obtained when the SOI leads the interannual LOD by 1 month; Eubanks et al. (1986) reported a maximum negative correlation coefficient during 1962–84 of about −0.5 when the SOI leads the interannual LOD by 3 months; Chao (1988) reported a maximum negative value for the correlation coefficient during 1972–86 of −0.68 for a 2-month lead time; and Dickey et al. (1993a) reported a maximum negative correlation coefficient during 1964–89 of −0.67 for a 1-month lead time. Since thermal winds are largely responsible for interannual LOD variations, Dickey et al. (2007), following Su et al. (2005), showed that the 1–2-month lag of interannual LOD variations behind the SOI can be explained by the 1–2-month lag of the tropical atmospheric temperature gradient behind the Niño-3.4 sea surface temperature index (another ENSO index that is highly correlated with the SOI).

In a detailed study of the 1982–83 ENSO event, Dickey et al. (1994b) showed that up to 92% of the observed interannual LOD variance could be explained by atmospheric wind and pressure fluctuations. They suggested that variations in oceanic angular momentum could explain the remaining signal. But studies (Gross et al., 2004; Johnson et al., 1999; Ponte et al., 2002) of the effects of oceanic processes show that they are only marginally effective in causing interannual LOD variations. For example, Gross et al. (2004) found that during 1980–2000, atmospheric winds were the dominant mechanism causing the length of day to change on interannual timescales, explaining 85.8% of the observed variance and having a correlation coefficient of 0.93 with the observations. The effect of atmospheric surface pressure changes explained only 2.6% of the observed variance and was not significantly correlated with the observations. However, including the effect of surface pressure changes with that of the winds increased the observed variance explained from 85.8% to 87.3%. Oceanic currents and bottom pressure changes were found to have only a marginal effect on interannual LOD variations, each explaining less than 1% of the observed variance and neither being significantly correlated with the observations. However, including their effects with those of atmospheric winds and surface pressure changes increased the observed variance explained from 87.3% to 87.9% and increased the correlation coefficient with the observations from 0.93 to 0.94.

The interannual LOD signal that remains after atmospheric and oceanic effects are removed may be caused by hydrologic processes (Chen, 2005; Chen et al., 2000a). For example, Chen (2005) found a significant correlation between the sum of the mass-balanced oceanic and hydrologic angular momenta and the observed interannual LOD variations from which atmospheric effects have been removed, although the modeled variations are of smaller amplitude than those observed. Like seasonal variations, better atmospheric, oceanic, and hydrologic models are needed to close the LOD budget on interannual timescales.

There is also a persistent oscillation in LOD over the last century with a period of about 6 years and an amplitude of about 0.12 ms that is not caused by atmospheric or oceanic fluctuations (Abarca del Rio et al., 2000; Djurovic and Paquet, 1996; Vondrák, 1977). Mound and Buffett (2003, 2006) suggested that it is the signature of a free mode of oscillation caused by the exchange of angular momentum between the mantle and core arising from gravitational coupling between the mantle and inner core. On the other hand, Gillet et al. (2010) suggested that it is the result of fast torsional waves in the core similar to the torsional oscillations that have been discussed as being the cause of decadal LOD variations (**Section 3.09.4.1.2**) but that now occur at much higher frequencies.

3.09.4.1.6 Intraseasonal variations and the Madden–Julian oscillation

Like the seasonal and interannual variations in the length of day, variations on intraseasonal timescales are also predominantly caused by changes in the angular momentum of the zonal winds (e.g., Dickey et al., 1992b, 2010; Eubanks, 1993; Hide and Dickey, 1991; Lambeck and Hopgood, 1981; Neef and Matthes, 2012; Rosen, 1993; Rosen et al., 1990; Yu et al., 1999; Zhou et al., 2008). The Madden–Julian oscillation (Madden and Julian, 1971, 1972, 1994, 2005) with a period of 30–60 days is the most prominent feature in the atmosphere on these timescales and a number of studies have shown that fluctuations in the zonal winds associated with this oscillation cause the length of day to change (Anderson and Rosen, 1983; Chao and Salstein, 2005; Dickey et al., 1991; Feissel and Gambis, 1980; Feissel and Nitschelm, 1985; Hendon, 1995; Langley et al., 1981b; Madden, 1987; Marcus et al., 2001).

Studies of the effects of oceanic processes show that they are only marginally effective in causing intraseasonal LOD variations (Gross et al., 2004; Johnson et al., 1999; Kouba and Vondrák, 2005; Ponte, 1997; Ponte and Ali, 2002; Ponte and Stammer, 2000). For example, Gross et al. (2004) found that during 1992–2000, atmospheric winds were the dominant mechanism causing the length of day to change on intraseasonal timescales, explaining 85.9% of the observed intraseasonal variance and having a correlation coefficient of 0.93 with the observations. Atmospheric surface pressure, oceanic currents, and ocean-bottom pressure were found to have only a minor effect on intraseasonal LOD changes, each explaining only about 3–4% of the observed variance. However, including the effect of surface pressure changes with that of the winds increased the variance explained from 85.9% to 90.2% and increased the correlation coefficient with the observations from 0.93 to 0.95. Additionally including the effects of changes in oceanic currents and bottom pressure further increased the variance explained from 90.2% to 92.2% and further increased the correlation coefficient from 0.95 to 0.96. Thus, although the impact of the oceans is relatively minor, closer agreement with the observations in the intraseasonal frequency band is obtained when the effects of oceanic processes are added to that of atmospheric.

While the overall contribution of the oceans to exciting intraseasonal LOD variations may be relatively minor, the effect on LOD of specific rapid changes in the oceans' circulation is detectable. For example, after accounting for the effect of the atmosphere, Marcus et al. (2012) showed that rapid changes in the LOD residual that occurred during 8–21 November 2009 were caused by rapid changes in the Antarctic Circumpolar Current as diagnosed from a data assimilating ocean model.

Hydrologic effects on intraseasonal LOD variations are thought to be relatively insignificant (Chen, 2005; Chen et al., 2000a).

3.09.4.2 Polar Motion

Observations of polar motion (**Figure 6**) show that it consists mainly of (1) a forced annual wobble having a nearly constant amplitude of about 100 mas, (2) the free Chandler wobble having a period of about 433 days and a variable amplitude ranging from about 100 to 200 mas, (3) quasiperiodic variations on decadal timescales having amplitudes of about 30 mas known as the Markowitz wobble, (4) a linear trend having a rate of about 3.5 mas year^{-1} and a direction toward 79°W longitude, and (5) smaller amplitude variations occurring on all measurable timescales. Here, the polar motion variations that are observed and the models that have been developed to explain them are reviewed.

The motion of the pole is usually described by giving the x- and y-components of its location in the terrestrial reference frame. But since polar motion is inherently a two-dimensional quantity, periodic motion of the pole can also be described by giving the amplitude A and phase α of its prograde and retrograde components defined by

$$\mathbf{p}(t) = p_x(t) - ip_y(t) = A_p e^{i\alpha_p} e^{i\sigma(t-t_o)} + A_r e^{i\alpha_r} e^{-i\sigma(t-t_o)} \qquad [68]$$

where the subscript p denotes prograde, the subscript r denotes retrograde, σ is the strictly positive frequency of motion, and t_o is the reference date. Prograde motion of the pole is circular

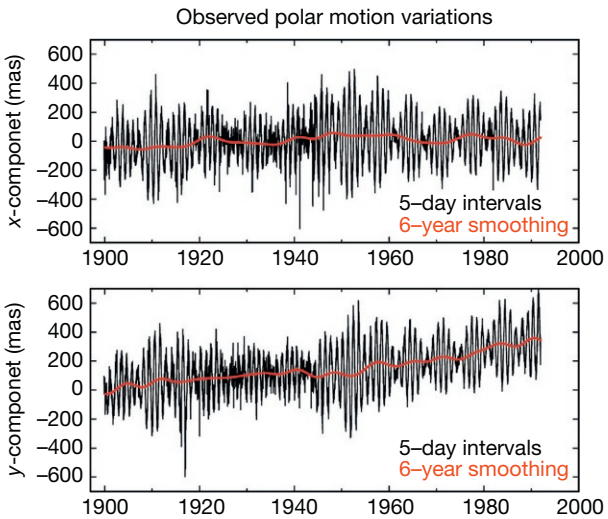


Figure 6 Observed polar motion variations from the Hipparcos optical astrometric series. The low-frequency variations shown in red were obtained by applying to the 5-day Hipparcos series a low-pass boxcar filter having a 6-year cutoff period. The beating between the 12-month annual wobble and the 14-month Chandler wobble is readily apparent.

motion in a counterclockwise direction; retrograde motion is circular motion in a clockwise direction. In general, the sum of prograde and retrograde circular motion is elliptical motion, with linear motion resulting when the amplitudes of the prograde and retrograde components are the same.

By eqns [41], [42], and [57], the observed polar motion variations are excited by changes in the equatorial components $h_x(t)$ and $h_y(t)$ of relative angular momentum and by changes in the equatorial components $\Omega\Delta I_{xz}(t)$ and $\Omega\Delta I_{yz}(t)$ of angular momentum due to changes in the mass distribution of the Earth. Like modeling the observed LOD changes, modeling the observed polar motion excitation requires computing both types of changes in angular momentum for the different components of the Earth system. From eqn [65] and again assuming for this purpose that the terrestrial reference frame is uniformly rotating at the rate Ω about the z-coordinate axis, the equatorial components of the mass term of the angular momentum can be written as

$$\Omega\Delta I_{xz}(t) = -\Omega\int_V \rho(\mathbf{r},t)r^2\sin\phi\cos\phi\cos\lambda\,\mathrm{d}V \qquad [69]$$

$$\Omega\Delta I_{yz}(t) = -\Omega\int_V \rho(\mathbf{r},t)r^2\sin\phi\cos\phi\sin\lambda\,\mathrm{d}V \qquad [70]$$

where ϕ is north latitude and λ is east longitude. Similarly, the equatorial components of the motion term of the angular momentum can be written as

$$h_x(t) = \int_V \rho(\mathbf{r},t)[r\sin\lambda v(\mathbf{r},t) - r\sin\phi\cos\lambda u(\mathbf{r},t)]\,\mathrm{d}V \qquad [71]$$

$$h_y(t) = \int_V \rho(\mathbf{r},t)[-r\cos\lambda v(\mathbf{r},t) - r\sin\phi\sin\lambda u(\mathbf{r},t)]\,\mathrm{d}V \qquad [72]$$

where $u(\mathbf{r},t)$ is the eastward component of the velocity and $v(\mathbf{r},t)$ is the northward component.

3.09.4.2.1 True polar wander and glacial isostatic adjustment

Determining an unbiased estimate of the linear trend in the observed path of the pole over the past 100 years is complicated by the presence of the annual, Chandler, and Markowitz wobbles. Various approaches have been taken to estimate the trend in the presence of these large-amplitude periodic and quasiperiodic variations (see **Table 8** for the resulting trend estimates). The annual wobble has been removed both by a least-squares fit (McCarthy and Luzum, 1996; Schuh et al., 2000, 2001; Wilson and Gabay, 1981) and by a seasonal adjustment of the polar motion series (Wilson and Vicente, 1980). The Chandler wobble has been removed both by a least-squares fit for periodic terms (Dickman, 1981; McCarthy and Luzum, 1996; Schuh et al., 2000, 2001) and by deconvolution (Vicente and Wilson, 2002; Wilson and Gabay, 1981; Wilson and Vicente, 1980). Smoothing has also been used to remove the annual and Chandler wobbles (Höpfner, 2004; Okamoto and Kikuchi, 1983). The decadal-scale variations have been removed by modeling them as being strictly periodic at a single frequency of 1/31 cpy and then least-squares fitting a sinusoid at this single frequency (Dickman, 1981; McCarthy and Luzum, 1996).

Rather than modeling the decadal variations as being strictly periodic at a single frequency, Gross and Vondrák (1999) devised a method to account for their quasiperiodic nature when estimating the linear trend. The annual and Chandler wobbles were first removed from the observations by applying to the polar motion series a low-pass boxcar filter having a 6-year cutoff period. A spectrum of the resulting low-pass filtered polar motion series was then computed and the linear trend was estimated by a simultaneous weighted least-squares fit for a mean, a trend, and periodic terms at the frequencies of all the peaks evident in the spectrum. The estimates for the linear trend that they obtained by applying this technique to the ILS, Hipparcos, and SPACE96 polar motion series are given in **Table 8**. **Figure 7** shows the observed low-pass filtered polar motion series they used (solid red lines); the modeled series they obtained by the simultaneous weighted

Table 8 Observed linear trend in the path of the pole

Rate (mas year^{-1})	*Direction (°W)*	*Data span*	*Source*
Hipparcos polar motion			
3.39	78.5	1899.7–1992.0	(a)
3.51 ± 0.01	**79.2 ± 0.20**	**1900.0–1992.0**	**(b)**
3.31 ± 0.05	76.1 ± 0.80	1899.7–1992.0	(c)
Hipparcos polar motion excitation			
2.84	73.03	1899–1992	(d)
ILS polar motion			
3.4	78	1900–1977	(e)
3.521 ± 0.094	80.1 ± 1.6	1899.8–1979.0	(f)
3.52	79.4	1899.8–1979.0	(g)
3.456	80.56	1899.0–1979.0	(h)
3.81 ± 0.07	75.5 ± 1.0	1899.8–1979.0	(b)
ILS polar motion excitation			
3.4	66	1900–1977	(e)
3.3	65	1901–1970	(i)
3.49	79.5	1899–1979	(d)
Latitude observations			
3.62	89	1900–1978	(j*)
3.51	79	1900–1978	(j†)
3.24	84.9	1899.8–1979.0	(k†)
2.97	77.7	1899.8–1979.0	(k‡)
Space-geodetic polar motion			
3.39 ± 0.53	85.4 ± 4.0	1976–1994	(l)
4.123 ± 0.002	73.9 ± 0.03	1976.7–1997.1	(b)
4.5 ± 0.1	68 ± 8	1976–1992	(m)
1.8 ± 0.4	58 ± 9	1992–2008	(m)
Combined astrometric and space-geodetic polar motion			
3.29	78.2	1900.0–1984.0	(n)
3.33 ± 0.08	75.0 ± 1.1	1899.8–1994.1	(l)
3.901 ± 0.022	65.17 ± 0.22	1891.0–1999.0	(o)
Combined astrometric and space-geodetic excitation			
3.35	76.3	1900–1999	(d)
3.54	69.92	1900–1999	(d)

The recommended estimate is given in bold.

Sources: (a) Vondrák et al. (1998), (b) Gross and Vondrák (1999), (c) Schuh et al. (2000, 2001), (d) Vicente and Wilson (2002), (e) Wilson and Vicente (1980), (f) Dickman (1981), (g) Chao (1983), (h) Okamoto and Kikuchi (1983), (i) Wilson and Gabay (1981), (j*) Zhao and Dong (1988) based on measurements taken at nine latitude observing stations, (j†) Zhao and Dong (1988) based on measurements taken at the five ILS observing stations located at Mizusawa, Kitab, Carloforte, Gaithersburg, and Ukiah, (k†) Vondrák (1994) based on measurements taken at the five ILS latitude observing stations located at Mizusawa, Kitab, Carloforte, Gaithersburg, and Ukiah, (k‡) Vondrák (1994) based on measurements taken at the four ILS latitude observing stations located at Mizusawa, Kitab, Carloforte, and Gaithersburg, (l) McCarthy and Luzum (1996), (m) Roy and Peltier (2011), (n) Vondrák (1985), (o) Höpfner (2004).

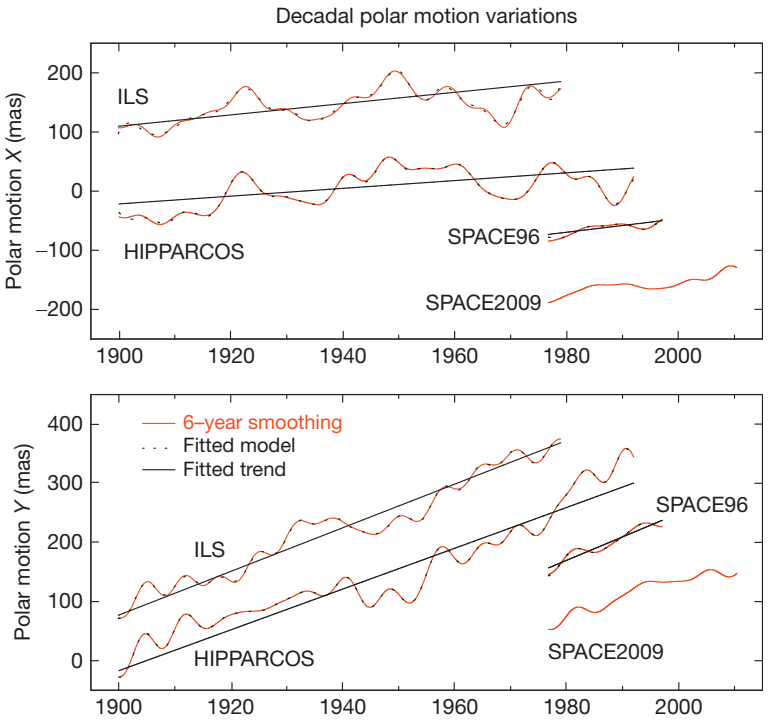


Figure 7 Observed decadal-scale polar motion variations (solid red lines) from the ILS optical astrometric series, the Hipparcos optical astrometric series, and the SPACE96 and SPACE2009 combined space-geodetic series. The decadal-scale variations were obtained by applying to the original series a low-pass boxcar filter having a 6-year cutoff period. The solid black lines show the linear trends in the pole path that were estimated from the ILS, Hipparcos, and SPACE96 series by Gross and Vondrák (1999). The dotted black lines show the model of the decadal variations that they used when estimating the trend. For clarity of display, the curves have been offset from each other by an arbitrary amount. Adapted from Gross RS and Vondrák J (1999) Astrometric and space-geodetic observations of polar wander. *Geophysical Research Letters* 26(14): 2085–2088.

least-squares fit for a mean, a trend, and all periodic terms evident in the spectrum (dotted black lines); and the resulting linear trend estimates (solid black lines). As can be seen, their model is an excellent fit to the observations, and hence, their estimates for the linear trend should be unbiased by the presence of the quasiperiodic decadal-scale polar motion variations. For this reason and because the Hipparcos series spans a greater length of time than the other series that they studied, the recommended estimate for the linear trend in the pole path is the estimate that they obtained for the Hipparcos series, namely, a trend of rate 3.51 ± 0.01 mas year^{-1} toward $79.2 \pm 0.20°$W longitude.

The estimates given in **Table 8** for the observed linear trend in the path of the pole are with respect to a terrestrial reference frame that has been attached to the mean lithosphere in such a manner that within it, the tectonic plates exhibit no net rotation. Argus and Gross (2004) argued that ideally, the motion of the pole should be given with respect to a reference frame that is attached to the mean solid Earth. They argued that it would therefore be better to use a reference frame that is attached to hotspots rather than the mean lithosphere because hotspots move slower than the mean lithosphere with respect to the mean solid Earth and hence better represent the mean solid Earth. Transforming the Hipparcos trend results of Gross and Vondrák (1999) to a hotspot reference frame, they found that during the past century, the linear trend in the path of the pole relative to hotspots was 4.03 mas year^{-1} toward 68.4° W longitude or about 15% faster and in a more eastward direction than the trend relative to the mean lithosphere.

One of the most important mechanisms acting to cause a linear trend in the path of the pole on timescales of a few thousand years is GIA. The isostatic adjustment of the solid Earth as it responds to the decreasing load on it following the last deglaciation causes the figure of the Earth to change and hence the pole to drift. Models of GIA show that its effect on the pole path is sensitive to the assumed value of lower mantle viscosity, to the assumed thickness and rheology of the lithosphere, to the treatment of the density discontinuity at 670 km depth, and to the assumed compressibility of the Earth model (e.g., Cambiotti et al., 2010; Johnston and Lambeck, 1999; Matsuyama et al., 2010; Mitrovica and Milne, 1998; Mitrovica and Wahr, 2011; Mitrovica et al., 2005, 2006, 2007; Nakada, 2000, 2002, 2009b; Peltier, 2007; Peltier and Drummond, 2010; Peltier and Jiang, 1996; Peltier and Luthcke, 2009; Peltier and Wu, 1983; Sabadini and Peltier, 1981; Sabadini and Vermeersen, 2002, 2004; Sabadini et al., 1982, 2002; Tamisea et al., 2002; Tsai and Stevenson, 2007; Vermeersen and Sabadini, 1996, 1999; Vermeersen et al., 1996, 1997, 1998; Wu and Peltier, 1984; Yuen et al., 1982).

However, GIA is not the only mechanism that will cause a trend in the pole path. The present-day change in glacier and ice sheet mass and the accompanying change in nonsteric sea level will also cause a linear trend in polar motion (Gasperini et al., 1986; James and Ivins, 1995, 1997; Mitrovica et al., 2006;

Nakada and Okuno, 2003; Peltier, 1988; Trupin, 1993; Trupin et al., 1992). But the effect of this mechanism is very sensitive to the still unknown present-day mass change of glaciers and ice sheets, particularly of the Antarctic ice sheet. By adopting various scenarios for the mass change of Antarctica, models predict that its mass change alone should cause a linear trend in the pole path ranging anywhere in rate from 0.31 to 4.46 mas year^{-1} and in direction from 101°W to 281°W longitude (James and Ivins, 1997).

Other mechanisms that may cause a linear trend in the path of the pole over the last or next century include continental drift (Dickman, 1979; Nakada, 2007, 2008), tectonic processes taking place under nonisostatic conditions (Sabadini and Vermeersen, 2004; Vermeersen and Vlaar, 1993; Vermeersen et al., 1994), plate subduction (Alfonsi and Spada, 1998; Greff-Lefftz, 2011; Ricard and Sabadini, 1990; Ricard et al., 1992, 1993; Richards et al., 1997; Rouby et al., 2010; Spada et al., 1992; Steinberger and Torsvik, 2010), mantle convection (Cambiotti et al., 2011; Chan et al., 2011a, 2011b; Greff-Lefftz, 2011; Nakada, 2008, 2009b; Richards et al., 1999; Schaber et al., 2009; Steinberger and O'Connell, 1997), upwelling mantle plumes (Greff-Lefftz, 2004, 2011; Rouby et al., 2010; Steinberger and O'Connell, 2002), earthquakes (Alfonsi and Spada, 1998; Chao and Gross, 1987; Chao et al., 1996b; Gross and Chao, 2006; Soldati and Spada, 1999; Spada, 1997; Zhou et al., 2013), and climate change (Landerer et al., 2009; Ponte et al., 2002).

As noted by Roy and Peltier (2011), the rate and direction of the linear trend in the pole path appear to have changed starting about 1992 (see **Table 8**). They speculated that this change may be caused by present-day changes in glacier and ice sheet mass, although changes in continental water storage may also be important. However, given the known presence of decadal-scale variations in the pole path, it is uncertain whether the pole will continue along this new path in the future or whether it will revert back to its pre-1992 long-term average rate and direction.

3.09.4.2.2 Decadal variations, the Markowitz wobble, and core–mantle interactions

By analyzing ILS measurements taken during 1900–59, Markowitz (1960, 1961) found that at long periods, the motion of the pole with respect to the Earth's crust and mantle includes a periodic component superimposed on a linear drift. He found that the periodic component, now known as the Markowitz wobble in his honor, had a period of 24 years and an amplitude of 22 mas and that the linear drift had a rate of 3.2 mas year^{-1} and a direction toward 60° W longitude, remarkably close to the recent determination of 3.51 mas year^{-1} toward 79.2° W longitude (Gross and Vondrák, 1999) that is recommended here.

In more recent rereductions of the ILS and other optical astrometric measurements, which extend more than three decades past those analyzed by Markowitz (1960, 1961), rather than appearing as a strictly periodic phenomenon of well-defined frequency, the Markowitz wobble now appears as a quasiperiodic variation on decadal timescales having an amplitude of about 30 mas (see **Figure 7**). However, since optical astrometric measurements are known to be corrupted by systematic errors, there has always been some doubt about the reality of the decadal variations seen in the ILS and Hipparcos series. Since the highly accurate space-geodetic measurements are less susceptible to systematic error than are optical astrometric measurements, any decadal variations seen in the space-geodetic measurements can be considered to be reliable. **Figure 7** compares the decadal variations seen in the SPACE96 and SPACE2009 combined space-geodetic series with those seen in the ILS and Hipparcos optical astrometric series. While decadal variations are seen to be present in the SPACE96 and SPACE2009 series and hence can be considered to be real, they have a smaller amplitude and a different phase than those seen in either the ILS or Hipparcos series. Thus, space-geodetic measurements both confirm the reality of decadal polar motion variations and demonstrate that the decadal variations seen in the ILS and Hipparcos series are unreliable and should be used only to place an upper bound on the size of the decadal variations.

The cause of the decadal-scale polar motion variations is currently unknown. Gross et al. (2005) found that redistribution of mass within the atmosphere and oceans cannot be the main excitation source of decadal polar motion variations during 1949–2002 since it amounts to only 20% (x-component) and 38% (y-component) of that observed and with the modeled excitation being 180° out of phase with that observed. However, the ocean model used in their study was not forced by mass changes associated with precipitation, evaporation, or runoff from rivers including that from glaciers and ice sheets and so had a constant total mass. Thus, their study did not address the question of the excitation of decadal polar motion by processes that change the total mass of the oceans, such as a nonsteric sea level height change associated with glacier and ice sheet mass change.

Wilson (1993) noted that an oscillation in global sea level on decadal timescales would excite decadal polar motion variations with a polarization similar to that observed, implying that mass change of the oceans is responsible for exciting the observed decadal polar motions. In fact, using a climate model, Celaya et al. (1999) showed that changes in Antarctic snow pack are capable of inducing decadal polar motion variations of nearly the same amplitude as that observed. But realistic estimates of mass change in glaciers and the Antarctic and other ice sheets, along with estimates of the accompanying nonsteric change in sea level, are required to further evaluate this possible source of decadal polar motion variations.

Changes in continental water storage may also be an important excitation source of decadal polar motion variations. But as with changes in present-day glacier and ice sheet mass, realistic estimates of changes in continental water storage are needed before this possible source of decadal polar motion variations can be accurately evaluated.

Since core–mantle processes are known to cause decadal variations in the length of day, they may also excite decadal variations in polar motion. But electromagnetic coupling between the core and mantle appears to be two to three orders of magnitude too weak (Greff-Lefftz and Legros, 1995) and topographic coupling appears to be too weak by a factor of three to ten (Asari et al., 2006; Dumberry, 2010; Greff-Lefftz and Legros, 1995; Hide et al., 1996; Hulot et al., 1996). In addition, the modeled decadal polar motion variations resulting from these studies show little agreement in phase with the observed variations.

Following the study of Jochmann (1989), Greiner-Mai et al. (2000, 2003) and Greiner-Mai and Barthelmes (2001) suggested that irregular motion of a tilted oblate inner core may excite decadal polar motion variations, although they did not account for the dynamic effects of the fluid core in their model. Dumberry and Bloxham (2002), who did account for the dynamic effects of the fluid core, suggested that a tilt of the inner core with respect to the mantle of only 0.07°, perhaps caused by an electromagnetic torque acting on the inner core, would generate gravitational and pressure torques on the mantle strong enough to excite decadal polar motion variations having amplitudes as large as those observed and, like the observed variations, having a preferred direction of motion. But when Mound (2005) examined this mechanism, he concluded that while torsional oscillations may excite both decadal LOD and polar motion variations, when the oscillations are constrained to match the observed LOD amplitude, the resulting electromagnetic torque on the inner core is too weak to excite decadal polar motions to their observed level.

Dumberry (2008) further studied the influence of the inner core on polar motion by examining the gravitational coupling between the inner core and density heterogeneities in the mantle. He found that his model is capable of producing decadal polar motions having amplitudes as large as those observed and that are polarized about a longitudinal axis. But when applying constraints to his model based on observed LOD variations, the amplitudes are found to be reduced to less than 1 mas, much less than those observed. While tantalizingly close results are obtained when the inner core is included in models of the effect of core–mantle processes on polar motion, such models have not yet explained all of the observed properties of the Markowitz wobble.

In closing, it is worth noting that the interaction between the core flow and a realistic nonaxial, nondipolar magnetic field will produce both axial and equatorial torques on the inner core and mantle, leading to LOD and polar motion variations that are coupled (Dumberry and Bloxham, 2002; Mound, 2005; Nakada, 2006). While such coupled motions have not yet been observed, it is worth searching the observations for them since it would provide additional insight into core–mantle processes.

3.09.4.2.3 Tidal wobbles and oceanic and atmospheric tides

Tidally induced deformations of the solid Earth caused by the second-degree zonal tide raising potential cause long-period changes in the Earth's rate of rotation (see **Section 3.09.4.1.3**). But since this potential is symmetrical about the polar axis, tidal deformations of the axisymmetric solid Earth cannot excite polar motion. However, due to the nonaxisymmetric shape of the coastlines, the second-degree zonal tide raising potential acting on the oceans can generate polar motion via the exchange of nonaxial oceanic tidal angular momentum with the solid Earth. Yoder et al. (1981) discussed the possible existence of long-period tidal variations in polar motion and tables of such variations predicted from theoretical ocean tide models have been given by Seiler (1990, 1991), Dickman (1993), Gross (1993), Brosche and Wünsch (1994), Seiler and Wünsch (1995), Weis (2006), and Dickman and Gross (2010).

Table 9 (also see Petit and Luzum, 2010, Table 8.4) gives the results obtained by Dickman and Gross (2010) from the theoretical ocean tide model of Dickman (1993) as revised by Dickman and Nam (1995). These results were shown by Dickman and Gross (2010) to fit the observed long-period tidal variations in polar motion better than any of the other models that they studied.

The results of Dickman and Gross (2010) for the effect of long-period ocean tides on polar motion were obtained from a theoretical ocean tide model that was not constrained by any type of data. As with UT1 and LOD, more accurate results for the effect of long-period ocean tides on polar motion can be expected to be obtained from tide models that are constrained

Table 9 Modeled variations in polar motion and polar motion excitation caused by long-period ocean tides

Tide	Argument					Period (days)	Polar motion				Polar motion excitation			
							Prograde		Retrograde		Prograde		Retrograde	
	l	l′	F	D	Ω		Amp (μas)	Phase (°)	Amp (μas)	Phase (°)	Amp (μas)	Phase (°)	Amp (μas)	Phase (°)
$M_{t'}$	1	0	2	0	1	9.12	4.43	−112.62	5.57	21.33	205.83	67.21	269.95	21.17
M_t	1	0	2	0	2	9.13	10.72	−112.56	13.48	21.30	497.59	67.27	652.59	21.14
$M_{f'}$	0	0	2	0	1	13.63	27.35	−91.42	30.59	13.31	841.32	88.42	1002.12	13.15
M_f	0	0	2	0	2	13.66	66.09	−91.31	73.86	13.27	2028.73	88.53	2414.94	13.11
M_{Sf}	0	0	0	2	0	14.77	5.94	−87.13	6.42	11.75	168.13	92.70	194.74	11.60
M_m	1	0	0	0	0	27.56	43.74	−56.70	31.12	−0.91	643.61	123.13	520.16	−1.06
M_{Sm}	−1	0	0	2	0	31.81	8.85	−51.11	5.42	−4.21	111.62	128.72	79.23	−4.36
S_{sa}	0	0	2	−2	2	182.62	86.48	−20.30	99.77	175.57	118.56	159.42	336.32	175.46
S_a	0	1	0	0	0	365.26	17.96	−17.38	152.15	170.60	3.33	161.60	332.53	170.51
M_n	0	0	0	0	1	−6798.38	208.17	166.89	186.98	166.67	221.43	166.88	175.07	166.68

l, l', F, D, and Ω are the Delaunay arguments, expressions for which are given in Simon et al. (1994). The period, given in solar days, is the approximate period of the term as viewed in the terrestrial reference frame. The amplitude (amp) and phase of the prograde and retrograde components of polar motion are defined by eqn [68]. The amplitude and phase of the prograde and retrograde components of polar motion excitation are similarly defined but with $\chi(t) = \chi_x(t) + i\chi_y(t)$.

Source: Dickman SR and Gross RS (2010) Rotational evaluation of a long-period spherical harmonic ocean tide model. *Journal of Geodesy* 84: 457–464, http://dx.doi.org/10.1007/s00190-010-0383-5.

by altimetric sea surface height data. Unfortunately, very few such data-constrained model results are available, and those that are available are either erroneous or incomplete. The results of Desai and Wahr (1995), as discussed by Gross et al. (1997), did not converge and are thus suspect even though data through T/P cycle 130 were used. And Ray and Egbert (2012) gave results for only the fortnightly ocean tide. More complete results for the effect of long-period ocean tides on polar motion based on altimetric-constrained ocean tide models are desirable.

Long-period tidal variations in polar motion were first observed by Chao (1994) and have been discussed by Gross et al. (1996b, 1997), Gross (2009b), and Ray and Egbert (2012). **Table 10** gives the results of Gross (2009b) for the observed variations that were determined by him by fitting periodic terms at the tabulated tidal frequencies to polar motion excitation observations spanning from 2 January 1980 to 8 September 2006 from which atmospheric and nontidal oceanic effects had been removed. While these results fully explain the observed long-period tidal variations during this time interval, they will most likely not be able to fully explain the observed variations outside this time interval.

Ocean tides in the diurnal and semidiurnal tidal bands also cause polar motion variations. While Yoder et al. (1981) discussed the possible existence of such polar motion variations, they were not actually observed until Dong and Herring (1990) detected them in VLBI measurements. Subdaily variations in polar motion at the diurnal and semidiurnal tidal frequencies have now been observed in measurements taken by VLBI (Artz et al., 2010, 2011; Böhm et al., 2012; Gipson, 1996; Haas and Wünsch, 2006; Herring, 1993; Herring and Dong, 1994; Sovers et al., 1993), SLR (Watkins and Eanes, 1994), GPS (Hefty et al., 2000; Rothacher et al., 2001; Steigenberger et al., 2006), and in combinations of GPS and VLBI measurements (Artz et al., 2012; Thaller et al., 2005, 2006, 2007). Predictions of subdaily tidal variations in polar motion using theoretical ocean tide models have been given by Seiler (1990, 1991), Dickman (1993), Gross (1993), Brosche and Wünsch (1994), Seiler and Wünsch (1995), and Weis (2006). Predictions of subdaily tidal variations in polar motion using tide models based upon T/P sea surface height measurements have been given by Chao et al. (1996a) and Chao and Ray (1997). Extensive tables of coefficients for the effect of diurnal and semidiurnal ocean tides on polar motion are given by Petit and Luzum (2010) (**Chapter 1.08**).

Comparisons of observations with models show the major role that ocean tides play in causing subdaily polar motion variations (**Figure 5(b)** and **5(c)**), with as much as 60% of the observed polar motion variance being explained by diurnal and semidiurnal ocean tides (Chao et al., 1996a; Thaller et al., 2006). Apart from errors in observations and models, the difference that remains (e.g., Schuh and Schmitz-Hübsch, 2000) may be due to nontidal atmospheric (Weber et al., 2002) and oceanic (Nastula et al., 2007a) effects, although Schindelegger et al. (2011) found that nontidal atmospheric effects can explain only about 25% of the residual polar motion signal.

The effects on polar motion of the radiational tides in the atmosphere have been studied by Zharov and Gambis (1996), Brzezinski et al. (2002a), and de Viron et al. (2005). However, since the oceans will respond dynamically to subdaily tidal variations in the atmospheric wind and pressure fields, the oceans will also contribute to the excitation of polar motion by the radiational tides. Brzezinski et al. (2004) predicted the effects on polar motion of the subdaily radiational tides in the atmosphere and oceans using a barotropic ocean model forced by the 6-h NCEP/NCAR reanalysis wind and pressure fields (also see Gross, 2005a). **Table 11** gives their results at prograde nearly diurnal frequencies. They found the largest effect to be at the prograde S_1 tidal frequency with an amplitude of 9 μas that is primarily excited by ocean-bottom pressure variations. Their predictions at retrograde nearly diurnal frequencies affect the nutations and are therefore not reproduced here. Since the NCEP/NCAR reanalysis wind and pressure fields are given every 6 h, the semidiurnal tidal frequency band is incompletely sampled, making it difficult to estimate atmospheric and oceanic effects on polar motion at these frequencies. More complete models for the effects of the radiational tides at nearly semidiurnal frequencies must await the availability of atmospheric fields sampled more often than every 6 h.

Table 10 Observed variations in polar motion and polar motion excitation caused by long-period ocean tides

Tide	Argument					Period (days)	Polar motion				Polar motion excitation			
							Prograde		Retrograde		Prograde		Retrograde	
	l	l′	F	D	Ω		Amp (μas)	Phase (°)	Amp (μas)	Phase (°)	Amp (μas)	Phase (°)	Amp (μas)	Phase (°)
$M_{t'}$	1	0	2	0	1	9.12	3.88	−128.58	1.65	−26.83	180.15	51.26	80.18	−26.98
M_t	1	0	2	0	2	9.13	9.36	−128.58	4.00	−26.83	434.55	51.26	193.41	−26.98
$M_{f'}$	0	0	2	0	1	13.63	28.62	−105.95	36.81	39.42	880.22	73.88	1205.93	39.27
M_f	0	0	2	0	2	13.66	69.18	−105.95	88.97	39.42	2123.44	73.88	2909.16	39.27
M_m	1	0	0	0	0	27.56	38.56	−111.41	83.24	34.77	567.37	68.42	1391.36	34.62

l, l', F, D, and Ω are the Delaunay arguments, expressions for which are given in Simon et al. (1994). The period, given in solar days, is the approximate period of the term as viewed in the terrestrial reference frame. The amplitude (amp) and phase of the prograde and retrograde components of polar motion are defined by eqn [68]. The amplitude and phase of the prograde and retrograde components of polar motion excitation are similarly defined but with $\boldsymbol{\chi}(t) = \chi_x(t) + i\chi_y(t)$.

Source: Gross RS (2009) An improved empirical model for the effect of long-period ocean tides on polar motion. *Journal of Geodesy* 83: 635–644, http://dx.doi.org/10.1007/s00190-008-0277-y.

Table 11 Modeled variations in polar motion caused by the diurnal radiational tide in the atmosphere and oceans

Tide	Fundamental argument						Period (solar days)	$p_x(t)$ (μas)		$p_y(t)$ (μas)	
	γ	l	l′	F	D	Ω		sin	cos	sin	cos
P_1	1	0	0	−2	2	−2	1.00275	0.1 ± 0.4	0.0 ± 0.4	−0.0 ± 0.4	0.1 ± 0.4
S_1	1	0	−1	0	0	0	1.00000	8.3 ± 0.4	−3.4 ± 0.4	3.4 ± 0.4	8.3 ± 0.4
K_1	1	0	0	0	0	0	0.99727	−0.1 ± 0.5	0.5 ± 0.5	−0.5 ± 0.5	−0.1 ± 0.5

γ is GMST reckoned from the lower culmination of the vernal equinox (GMST + π). l, l', F, D, and Ω are the Delaunay arguments, expressions for which are given in Simon et al. (1994). The period, given in solar days, is the approximate period of the term as viewed in the terrestrial reference frame. Since the diurnal radiational tide is seasonally modulated, it also affects polar motion at the P_1 and K_1 tidal frequencies.

Source: Brzezinski A, Ponte RM, and Ali AH (2004) Nontidal oceanic excitation of nutation and diurnal/semidiurnal polar motion revisited. *Journal of Geophysical Research* 109: B11407, http://dx.doi.org/10.1029/2004JB003054.

3.09.4.2.4 The Chandler wobble and its excitation

Any irregularly shaped solid body rotating about some axis that is not aligned with its figure axis will freely wobble as it rotates (Euler, 1765). The Eulerian free wobble of the Earth is known as the Chandler wobble in honor of Seth Carlo Chandler, Jr., who first observed it (Chandler, 1891). Unlike the forced wobbles of the Earth, such as the annual wobble, whose periods are the same as the periods of the forcing mechanisms, the period of the free Chandler wobble is a function of the internal structure and rheology of the Earth (see **Section 3.09.2.1**) and its decay time constant, or quality factor Q, is a function of the dissipation mechanisms acting to dampen it. The observed values for the period and Q of the Chandler wobble can therefore be used to better understand the internal structure of the Earth and the dissipation mechanisms, such as mantle anelasticity, that dampen the Chandler wobble causing its amplitude to decay in the absence of excitation.

Determining an unbiased estimate for the period and Q of the Chandler wobble is complicated by the relatively short duration of the observational record and by incomplete and inaccurate models of the mechanisms acting to excite it. In the absence of any knowledge of its excitation, statistical models of the excitation can be adopted (e.g., Jeffreys, 1972; Ooe, 1978; Wilson and Haubrich, 1976; Wilson and Vicente, 1980, 1990). Since atmospheric, oceanic, and hydrologic processes are thought to be major sources of Chandler excitation, its period and Q have also been estimated by using AAM data (Furuya and Chao, 1996; Kuehne et al., 1996), atmospheric and oceanic angular momentum data (Gross, 2005b; Seitz and Kutterer, 2005; Seitz et al., 2012), and atmospheric, oceanic, and hydrologic angular momentum data (Wilson and Chen, 2005). By fitting a small set of basic Earth parameters to observed nutation amplitudes and precession rate, semianalytic estimates of the period and Q of the Chandler wobble have been determined by Mathews et al. (2002); a recent update of the estimates has been given by Chen and Shen (2010). **Table 12** gives the resulting values for the period and Q of the Chandler wobble.

Gross (2005b) discussed the sensitivity of the estimated period and Q of the Chandler wobble to the length of the data sets that he analyzed, concluding that data sets spanning at least 31 years are needed to obtain stable estimates. He also noted the need for Monte Carlo simulations to determine corrections to the bias of the estimated Q values. Since Gross (2005b) did not do this, but since it was done by Wilson and Vicente (1990) who also used data spanning 86 years to estimate the period and Q of the Chandler wobble, the recommended estimate is that determined by them, namely, a period of 433.0 ± 1.1 (1σ) solar days and a Q of 179 with a 1σ range of 74–789.

Table 12 Estimated period and Q of the Chandler wobble

Period (solar days)	Q	Data span (years)	Source
Statistical excitation			
433.2 ± 2.2	63 (36, 192)	67.6	(a)
434.0 ± 2.6	100 (50, 400)	70	(b)
434.8 ± 2.0	96 (50, 300)	76	(c)
433.3 ± 3.1	170 (47, 1000)	78	(d)
433.0 ± 1.1	**179 (74, 789)**	**86**	**(e)**
433.1 ± 1.7	–	93	(f)
Atmospheric excitation			
439.5 ± 2.1	72 (30, 500)	8.6	(g)
433.7 ± 1.8	49 (35, 100)	10.8	(h)
430.8	41	10	(i)
Atmospheric and oceanic excitation			
429.4	107	10	(i)
431.9	83	51	(i)
432.98	97	60	(j)
Semianalytic			
430.3	88.4	20	(k)
433.03	100.20	20	(l)

The recommended estimate is given in bold. The 1σ confidence interval for the Q estimates is given in parentheses.

Sources: (a) Jeffreys (1972), (b) Wilson and Haubrich (1976), (c) Ooe (1978), (d) Wilson and Vicente (1980), (e) Wilson and Vicente (1990), (f) Vicente and Wilson (1997), (g) Kuehne et al. (1996), (h) Furuya and Chao (1996), (i) Gross (2005b), (j) Seitz et al. (2012), (k) Mathews et al. (2002), (l) Chen and Shen (2010).

With these recommended values for the period T and quality factor Q of the Chandler wobble, in the absence of excitation, it would freely decay to the minimum rotational energy state of rotation about the figure axis with an e-folding amplitude decay time constant $\tau = 2QT/2\pi$ of about 68 years. But a damping time of 68 years is short on a geologic timescale and since the amplitude of the Chandler wobble has at times been observed to actually increase, some mechanism or mechanisms must be acting to excite it. Since its discovery, many possible excitation mechanisms have been studied, including core–mantle interactions (Gire and Le Mouël, 1986; Hinderer et al., 1987, 1990; Jault and Le Mouël, 1993; Nakada, 2009a,

2011), earthquakes (Gross, 1986; Souriau and Cazenave, 1985), continental water storage (Brzezinski et al., 2012; Chao et al., 1987; Hinnov and Wilson, 1987; Jin et al., 2010; Kuehne and Wilson, 1991; Liao et al., 2004), atmospheric wind and surface pressure variations (Aoyama, 2005; Aoyama and Naito, 2001; Aoyama et al., 2003; Furuya et al., 1996, 1997; Liao et al., 2007; Stuck et al., 2005; Wahr, 1983; Wilson and Haubrich, 1976), and oceanic current and bottom pressure variations (Bizouard et al., 2011; Brzezinski and Nastula, 2002; Brzezinski et al., 2002b; Gross, 2000b; Gross et al., 2003; Liao, 2005; Liao et al., 2003; Nastula et al., 2012; Ponte, 2005; Ponte and Stammer, 1999; Seitz and Schmidt, 2005; Seitz et al., 2004, 2005; Thomas et al., 2005; Zotov and Bizouard, 2012).

While there is growing agreement that the Chandler wobble is excited by a combination of atmospheric, oceanic, and hydrologic processes, the relative contribution of each process to its excitation is still being debated. For example, Gross et al. (2003) studied the excitation of the Chandler wobble during 1980–2000, finding that atmospheric winds and surface pressure and oceanic currents and bottom pressure combined are significantly coherent with and have enough power to excite the Chandler wobble. They found that during this time interval, the observed power in the Chandler band is 1.90 mas^2, with the sum of all atmospheric and oceanic excitation processes having more than enough power, at 2.22 mas^2, to excite the Chandler wobble. Ocean-bottom pressure variations were found to be the single most effective process exciting the Chandler wobble, with atmospheric surface pressure variations having about 2/3 as much power as ocean-bottom pressure variations and with the power of winds and currents combined being less than 1/3 the power of the combined effects of surface and bottom pressure. The conclusion that oceanic processes are more effective than atmospheric in exciting the Chandler wobble has also been reached by Gross (2000b), Brzezinski and Nastula (2002), Brzezinski et al. (2002b, 2012), and Liao et al. (2003, 2004). Studies of the excitation of the Chandler wobble using coupled atmosphere–ocean climate models also confirm that atmospheric and oceanic processes have enough power to maintain the Chandler wobble and indicate that the relative contribution of individual atmospheric and oceanic processes changes with time (Celaya et al., 1999; Leuliette and Wahr, 2002; Ponte et al., 2002). However, other studies have concluded that atmospheric processes alone have enough power to excite the Chandler wobble (Aoyama, 2005; Aoyama and Naito, 2001; Aoyama et al., 2003; Furuya et al., 1996, 1997). The resolution of the debate about the relative importance of atmospheric, oceanic, and hydrologic processes to exciting the Chandler wobble awaits the availability of more accurate models of these processes and, because the power needed to maintain the Chandler wobble depends upon its damping (Gross, 2000b), of more accurate estimates of its period and Q.

3.09.4.2.5 Seasonal wobbles

While continuing to analyze variation of latitude measurements, Chandler (1892) soon discovered that the wobbling motion of the Earth includes an annual component in addition to the 14-month component. The annual wobble is a forced wobble of the Earth that is caused largely by the annual appearance of a high atmospheric pressure system over Siberia every winter (e.g., Eubanks, 1993; Lambeck, 1980, 1988; Munk and MacDonald, 1960).

Chao and Au (1991) showed that during 1980–88, the amplitude of the prograde annual polar motion excitation can be accounted for by atmospheric wind and pressure fluctuations, with equatorial winds contributing about 25% and pressure fluctuations contributing about 75% to the total atmospheric excitation. But even though the amplitude of the observed prograde annual polar motion excitation is consistent with atmospheric wind and pressure fluctuations, there is a rather large phase discrepancy of about 30°. Furthermore, no agreement was found by Chao and Au (1991) between the observed retrograde annual polar motion excitation and that due to atmospheric processes, with atmospheric excitation being about twice as large as the observed excitation. Discrepancies as large as a factor of two in amplitude were also found by Chao and Au (1991) between observed semiannual polar motion excitation and that due to atmospheric processes (also see Aoyama and Naito, 2000; Barnes et al., 1983; Chao, 1993; Dobslaw et al., 2010; King and Agnew, 1991; Kolaczek et al., 2003; Merriam, 1982; Nastula and Kolaczek, 2002; Nastula and Salstein, 2012; Nastula et al., 2005, 2009; Neef and Matthes, 2012; Stuck et al., 2005; Wahr, 1983; Wilson and Haubrich, 1976; Zhou et al., 2006, 2008).

Since the agreement with atmospheric excitation is poor except for the prograde annual amplitude, other processes must be contributing to the excitation of seasonal wobbles. Recently, near-global general circulation models of the oceans have been used to investigate the contribution that nontidal oceanic processes make to exciting the seasonal wobbles of the Earth (Brzezinski et al., 2005; Chen et al., 2004a; Dobslaw et al., 2010; Furuya and Hamano, 1998; Gross, 2004; Gross et al., 2003; Johnson, 2005; Johnson et al., 1999; Liu et al., 2007; Nastula et al., 2000, 2003, 2012; Ponte and Stammer, 1999; Ponte et al., 1998, 2001; Seitz and Schmidt, 2005; Seitz et al., 2004; Wünsch, 2000; Zhong et al., 2006; Zhou et al., 2005). These studies have shown that adding nontidal oceanic excitation to atmospheric improves the agreement with the observed excitation. For example, **Table 13** gives the results obtained by Gross et al. (2003) for the atmospheric excitation and oceanic excitation of polar motion at the annual and semiannual frequencies. They found that atmospheric processes were more effective than oceanic in exciting the annual and semiannual wobbles. Atmospheric surface pressure variations were found to be the single most important mechanism exciting the annual and semiannual wobbles, with the sum of surface and ocean-bottom pressure variations being about two to three times as effective as the sum of winds and currents.

A rather large residual remains after the effects of the atmosphere and oceans are removed from the observed seasonal polar motion excitation. This residual not only is probably at least partly due to errors in the atmospheric and oceanic models but also could be due to the neglect of other excitation processes such as hydrologic processes (Chao and O'Connor, 1988; Chen and Wilson, 2005; Chen et al., 2000a; Dobslaw et al., 2010; Hinnov and Wilson, 1987; Kuehne and Wilson, 1991; Nastula and Kolaczek, 2005; Nastula and Salstein, 2012; Wünsch, 2002). Wünsch (2002) summarized the contribution

Table 13 Observed and modeled nontidal polar motion excitation at annual and semiannual frequencies

Excitation process	*Annual*				*Semiannual*			
	Prograde		*Retrograde*		*Prograde*		*Retrograde*	
	Amplitude (mas)	*Phase (°)*	*Amplitude (mas)*	*Phase (°)*	*Amplitude (mas)*	*Phase (°)*	*Amplitude (mas)*	*Phase (°)*
Observed excitation	14.52 ± 0.33	−62.77 ± 1.30	7.53 ± 0.33	−120.09 ± 2.50	5.67 ± 0.33	107.56 ± 3.32	5.80 ± 0.33	123.66 ± 3.24
Atmospheric excitation								
Winds	2.97 ± 0.12	−34.92 ± 2.30	2.04 ± 0.12	12.06 ± 3.36	0.36 ± 0.12	71.76 ± 19.0	0.55 ± 0.12	−134.43 ± 12.4
Surface pressure (IB)	15.12 ± 0.17	−101.92 ± 0.66	15.05 ± 0.17	−105.30 ± 0.66	2.60 ± 0.17	47.45 ± 3.81	4.75 ± 0.17	103.77 ± 2.08
Winds and surface pressure	16.51 ± 0.23	−92.38 ± 0.81	14.23 ± 0.23	−98.00 ± 0.94	2.93 ± 0.23	50.36 ± 4.58	4.49 ± 0.23	109.77 ± 2.99
Oceanic excitation								
Currents	2.31 ± 0.09	39.72 ± 2.13	2.11 ± 0.09	50.70 ± 2.33	1.32 ± 0.09	176.16 ± 3.71	1.41 ± 0.09	−143.46 ± 3.49
Bottom pressure	3.45 ± 0.11	63.18 ± 1.87	3.42 ± 0.11	110.24 ± 1.89	0.77 ± 0.11	133.89 ± 8.41	1.48 ± 0.11	−136.46 ± 4.35
Currents and bottom pressure	5.64 ± 0.16	53.81 ± 1.61	4.84 ± 0.16	88.22 ± 1.87	1.96 ± 0.16	160.89 ± 4.62	2.89 ± 0.16	−139.87 ± 3.14
Atmospheric and oceanic excitation								
Winds and currents	4.22 ± 0.15	−3.09 ± 2.08	3.91 ± 0.15	31.71 ± 2.24	1.28 ± 0.15	160.38 ± 6.83	1.96 ± 0.15	−140.92 ± 4.48
Surface and bottom pressure	11.82 ± 0.20	−97.61 ± 0.99	12.43 ± 0.20	−114.51 ± 0.94	2.75 ± 0.20	63.61 ± 4.26	4.22 ± 0.20	121.55 ± 2.78
Total of all modeled excitation	12.23 ± 0.28	−77.50 ± 1.32	9.43 ± 0.28	−101.18 ± 1.71	2.90 ± 0.28	89.70 ± 5.57	4.41 ± 0.28	147.63 ± 3.66

IB, inverted barometer. The amplitude and phase of the prograde and retrograde components of polar motion excitation are defined by eqn [68] but with $\chi(t) = \chi_x(t) + i\chi_y(t)$. The reference date to the phase is 1 January 1990.

Source: Gross RS, Fukumori I, and Menemenlis D (2003) Atmospheric and oceanic excitation of the Earth's wobbles during 1980–2000. *Journal of Geophysical Research* 108(B8): 2370. http://dx.doi.org/10.1029/2002JB002143.

of soil moisture and snow load to exciting the annual and semiannual wobbles. Although the available soil moisture models exhibit large differences, Wünsch (2002) concluded that soil moisture and snow load effects are important contributors to exciting the annual and semiannual wobbles. Climate models have also been used to study atmospheric, oceanic, and hydrologic excitation of seasonal polar motion (Celaya et al., 1999; Neef and Matthes, 2012; Ponte et al., 2002; Zhong et al., 2003).

3.09.4.2.6 Nonseasonal wobbles

Like the seasonal wobbles, the wobbling motion of the Earth on interannual timescales is a forced response of the Earth to its excitation mechanisms. Abarca del Rio and Cazenave (1994) compared the observed excitation of the Earth's wobbles to atmospheric excitation during 1980–91, finding similar fluctuations both in components on timescales between 1 and 3 years and in the γ-component on timescales between 1.2 and 8 years, but only when the atmospheric excitation is computed assuming the oceans fully transmit the imposed atmospheric pressure variations to the floor of the oceans (rigid ocean approximation). Chao and Zhou (1999) studied the correlation of the observed polar motion excitation functions on interannual timescales during 1964–94 with the SOI and the North Atlantic Oscillation Index (NAOI). Although little agreement was found with the SOI, a significant agreement was found with the NAOI, especially for the x-component of the observed excitation function, indicating a possible meteorologic origin of the interannual wobbles (also see Nastula et al., 2009; Neef and Matthes, 2012; Zhou et al., 1998). The lack of an ENSO signal in polar motion, as evidenced by the lack of a correlation between the SOI and interannual polar motion variations found by Chao and Zhou (1999), was explained by Marcus et al. (2010) as being a consequence of the mitigation of atmospheric ellipsoidal torques by the geometric distribution of the land and oceans.

Johnson et al. (1999) compared the observed polar motion excitation functions on interannual timescales during 1988–98 with atmospheric and oceanic excitation functions, finding only weak agreement between the observed and modeled excitation functions. For periods between 1 year and 6 years excluding the annual cycle, oceanic processes were found by Gross et al. (2003) to be much more important than atmospheric in exciting interannual polar motions, explaining 33% of the observed variance compared with 6% for atmospheric processes and having a correlation coefficient of 0.59 with the observations compared with 0.28 for atmospheric processes. Ocean-bottom pressure variations were found to be the single most important excitation mechanism, explaining 30% of the observed variance and having a correlation coefficient of 0.59 with the observations. Although the other processes were not nearly as effective as ocean-bottom pressure variations in exciting interannual polar motions, when all the processes were combined, they were found to explain 40% of the observed variance and to have a correlation coefficient of 0.64 with the observations (also see Brzezinski et al., 2005; Chen and Wilson, 2005; Johnson, 2005).

Like the seasonal and interannual wobbles, the wobbling motion of the Earth on intraseasonal timescales is a forced response of the Earth to its excitation mechanisms. Eubanks et al. (1988) were the first to study the Earth's wobbles on timescales between 2 weeks and several months, concluding that these rapid polar motions during 1983.75–86.75 are at least partially driven by atmospheric surface pressure changes and suggesting that the remaining excitation may be caused by dynamic ocean-bottom pressure variations. Subsequent studies confirmed the importance of atmospheric processes in exciting rapid polar motions (di Leonardo and Dickman, 2004; Feldstein, 2008; Gross and Lindqwister, 1992; Kolaczek et al., 2000a,b; Kosek et al., 1995; Kuehne et al., 1993; Nastula, 1992, 1995, 1997; Nastula and Salstein, 1999; Nastula et al., 1990, 1997, 2009; Neef and Matthes, 2012; Salstein, 1993; Salstein and Rosen, 1989; Schuh and Schmitz-Hübsch, 2000; Stieglitz and Dickman, 1999; Zhou et al., 2006, 2008), although the existence of significant discrepancies indicates that nonatmospheric processes must also play an important role.

The contribution of oceanic processes to exciting rapid polar motions has been studied using both barotropic (Nastula and Ponte, 1999; Ponte, 1997) and baroclinic (Bizouard and Seoane, 2010; Brzezinski et al., 2005; Chen et al., 2004a; de Viron et al., 2004a; Gross et al., 2003; Johnson et al., 1999; Lambert et al., 2006; Nastula and Kolaczek, 2005; Nastula et al., 2000; Ponte et al., 1998, 2001; Zhou et al., 2005) models of the oceans. Such studies have shown that while better agreement with the observations is obtained when oceanic excitation is added to that of the atmosphere, significant discrepancies still remain. For example, Gross et al. (2003) found that on intraseasonal timescales, with periods between 5 days and 1 year excluding the seasonal cycles, atmospheric processes were found to be more effective than oceanic in exciting polar motion, explaining 45% of the observed variance compared with 19% for oceanic processes and having a correlation coefficient of 0.67 with the observations compared with 0.44 for oceanic processes. Of these processes, atmospheric surface pressure variations were found to be the single most effective process exciting intraseasonal polar motions, with winds being the next most important process; ocean-bottom pressure was about half as effective as atmospheric surface pressure, and currents were about half as effective as winds. Atmospheric winds and surface pressure and oceanic currents and bottom pressure combined explained 65% of the observed variance and had a correlation coefficient of 0.81 with the observations.

On the shortest intraseasonal timescales, Ponte and Ali (2002) used the NCEP/NCAR reanalysis atmospheric model and a barotropic ocean model forced by the NCEP/NCAR reanalysis surface pressure and wind stress fields to study atmospheric and oceanic excitation of polar motion from July 1996 to June 2000 on timescales of 2–20 days. They confirmed the importance of oceanic processes in exciting rapid polar motions, finding that about 60% of the observed polar motion excitation variance between periods of 6 and 20 days is explained by atmospheric excitation, increasing to about 80% when oceanic excitation is added to that of the atmosphere (see also Bizouard and Seoane, 2010; Johnson, 2008; Kouba, 2005; Lambert et al., 2006; Nastula and Kolaczek, 2007; Nastula et al., 2002).

A spectra of subdaily atmospheric and oceanic excitation functions (Brzezinski et al., 2002a; Gross, 2005a) reveal the presence of atmospheric normal modes: (1) a peak at

about −0.83 cycles per day (cpd) due to the ${\psi_1}^1$ atmospheric normal mode (Masaki, 2008; Sidorenkov, 2003); (2) a peak of smaller amplitude at about +1.8 cpd due to the ${\xi_2}^1$ atmospheric normal mode; and (3) slight enhancement in power at about −0.12 cpd due to the ${\psi_3}^1$ atmospheric normal mode (Brzezinski, 1987; Feldstein, 2006). Because of their small amplitudes, the effects of these normal modes on polar motion are difficult to detect, although the effect of the ψ_3^1 mode has been detected (Bizouard and Seoane, 2010; Eubanks et al., 1988; Feldstein, 2008).

Apart from errors in observations and models, the residual that remains on broad intraseasonal to interannual timescales after modeled atmospheric and oceanic effects have been removed from the observations may be caused by hydrologic processes (e.g., Fernández et al., 2007; Göttl and Seitz, 2009).

3.09.5 Mass Redistribution, Gravity, and Earth Rotation

Redistribution of mass within the Earth system causes the Earth's gravitational field to change and, by changing the Earth's inertia tensor, causes the Earth's rotation to change. Because the elements of the Earth's inertia tensor are related to the degree-2 coefficients of the Earth's gravitational field, measurements of these coefficients can be used to study changes in the Earth's rotation caused by mass redistribution. The relations between the degree-2 coefficients of the gravitational field and the elements of the inertia tensor are derived here and the results of using gravitational field measurements to study changes in the Earth's rotation are reviewed.

In geodesy, the gravitational potential $U(\mathbf{r_o})$ of the Earth evaluated at some external field point $\mathbf{r_o}(r_o, \theta_o, \phi_o)$ is usually given in the form (e.g., Kaula, 1966)

$$U(\mathbf{r_o}) = \frac{GM}{r_o}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\left(\frac{a}{r_o}\right)^l (C_{lm}\cos m\phi_o + S_{lm}\sin m\phi_o)\,\widetilde{P}_{lm}(\cos\theta_o) \quad [73]$$

where G is the gravitational constant, M is the mass of the Earth, a is some reference radius that is less than r_o and that is usually taken to be the radius of some spherically symmetrical Earth model, and $\widetilde{P}_{lm}$ are the normalized associated Legendre functions of degree l and order m. The dimensionless expansion coefficients C_{lm} and S_{lm}, known as Stokes coefficients, are the normalized multipole moments of the Earth's density field $\rho(\mathbf{r})$:

$$C_{lm} + iS_{lm} = \frac{N_{lm}}{Ma^l}\int_{V_o} r^l\, Y_{lm}(\theta, \phi)\,\rho(\mathbf{r})\,dV \quad [74]$$

where the integral extends over the volume V_o of the Earth, the Y_{lm} are the fully normalized surface spherical harmonic functions, and the N_{lm} are normalization factors (Chao and Gross, 1987):

$$N_{lm} = (-1)^m\frac{2}{2l+1}\sqrt{(2-\delta_{m0})\pi} \quad [75]$$

In Cartesian coordinates, the degree-2, order-0, and order-1 Stokes coefficients can be written as

$$C_{20} = \frac{1}{\sqrt{20}Ma^2}\int_{V_o}\left[2r^2 - 3\left(x^2+y^2\right)\right]\rho(\mathbf{r})\,dV \quad [76]$$

$$C_{21} + iS_{21} = \sqrt{\frac{3}{5}}\frac{1}{Ma^2}\int_{V_o}(xz + iyz)\,\rho(\mathbf{r})\,dV \quad [77]$$

The inertia tensor $\mathbf{I}$ is defined as

$$\mathbf{I} = \int_{V_o}\left(r^2\mathbf{I} - \mathbf{rr}\right)\rho(\mathbf{r})\,dV \quad [78]$$

where $\mathbf{I}$ is the identity tensor. In Cartesian coordinates, the elements of the inertia tensor that, when they change, affect the Earth's rotation (see eqns [38]–[43]) can be written as

$$I_{xz} = -\int_{V_o} xz\,\rho(\mathbf{r})\,dV \quad [79]$$

$$I_{yz} = -\int_{V_o} yz\,\rho(\mathbf{r})\,dV \quad [80]$$

$$I_{zz} = \int_{V_o}\left(x^2+y^2\right)\rho(\mathbf{r})\,dV \quad [81]$$

Thus, the desired relations between the degree-2 Stokes coefficients and those elements of the inertia tensor that affect the Earth's rotation are

$$C_{20} = \frac{1}{\sqrt{20}Ma^2}\left(T - 3I_{zz}\right) \quad [82]$$

$$C_{21} + iS_{21} = -\sqrt{\frac{3}{5}}\frac{1}{Ma^2}\left(I_{xz} + iI_{yz}\right) \quad [83]$$

where T is the trace of the inertia tensor:

$$T = I_{xx} + I_{yy} + I_{zz} = \int_{V_o} 2r^2\rho(\mathbf{r})\,dV \quad [84]$$

Rochester and Smylie (1974) had shown that the trace of the inertia tensor T does not change for many types of mass redistribution including those for which the total mass of the system does not change. For example, the trace of the inertia tensor does not change if the system consists of the entire atmospheric, oceanic, and land–hydrologic system because the mass of this entire system does not change as water in its various phases moves between the atmosphere, oceans, and land. However, if the system consists of just, say, the atmosphere, then the trace of the inertia tensor does change because the mass of the atmosphere changes as water evapotranspires into and precipitates out of the atmosphere.

Equations [82] and [83] are quite general. But applying them to processes that load the solid Earth is a bit cumbersome because the integrals in eqns [76], [77], and [79]–[81] extend over the entire volume of the Earth and thus include the deformation of the solid Earth caused by the loading process. That is, eqns [82] and [83] apply to the sum of the loading process and the solid Earth's response to the load. It is usually more convenient to separate these and account for the deformational response of the solid Earth to the load by using load Love numbers. In this case, eqn [74] for the degree-2 Stokes coefficients can be written as

$$C_{2m} + iS_{2m} = \left(1 + k'_2 + \Delta k'_{an}\right)\frac{N_{2m}}{Ma^2}\int_{V_l} r^2\, Y_{2m}(\theta, \phi)\,\rho(\mathbf{r})\,dV \quad [85]$$

where the integral now extends over just the volume V_l of the load, k'_2 is the second-degree load Love number, and $\Delta k'_{an}$

accounts for the effects of mantle anelasticity on the second-degree load Love number. Applying eqns [79]–[81] to just the load, the desired relations between time-dependent changes in the degree-2 Stokes coefficients and the relevant elements of the inertia tensor of the load are

$$\Delta C_{20} = \frac{1 + k_2' + \Delta k_{an}'}{\sqrt{20} M a^2} \left(\Delta T^l - 3\Delta \mathbf{I}_{zz}^l \right) \quad [86]$$

$$\Delta C_{21} + i\Delta S_{21} = -\left(1 + k_2' + \Delta k_{an}'\right) \frac{\sqrt{3/5}}{M a^2} \left(\Delta \mathbf{I}_{xz}^L + i\Delta \mathbf{I}_{yz}^L \right) \quad [87]$$

where the time dependence has been suppressed for brevity and the superscript l indicates that the quantity is for just the load, not the entire Earth.

Load Love numbers are also used to account for the effect on the Earth's rotation of the deformation of the solid Earth caused by surficial loading processes (see eqns [41]–[43]). Thus, as in eqns [86] and [87], the changes in the elements of the inertia tensor in eqns [41]–[43] are for just the load. By combining eqns [41]–[43] and [86] and [87] after setting the relative angular momenta terms $h_i(t)$ in eqns [41]–[43] to zero, the following relation between the Earth rotation excitation functions and the degree-2 Stokes coefficients can be derived:

$$\Delta C_{20}(t) = \frac{1 + k_2' + \Delta k_{an}'}{\sqrt{20} M a^2} \left[\Delta T^l(t) - \frac{3 C_m / 0.997}{1 + \alpha_3 \left(k_2' + \Delta k_{an}'\right)} \chi_z(t) \right] \quad [88]$$

$$\Delta C_{21}(t) + i\Delta S_{21}(t) = -\frac{\sqrt{3/5}}{M a^2} \frac{C - A'}{1.608} \left[\chi_x(t) + i\chi_y(t) \right] \quad [89]$$

Equations [88]–[89] can be used to relate observed changes in the degree-2 Stokes coefficients to changes in the Earth rotation excitation functions caused by redistribution of mass within the Earth's surface geophysical fluids. Observations of the time-dependent degree-2 Stokes coefficients are currently being provided by SLR and the Gravity Recovery and Climate Experiment (GRACE) twin satellite mission. The use of these gravitational field observations to study changes in the Earth's rotation is discussed in the succeeding text.

3.09.5.1 GRACE and SLR Measurements of Earth Rotation Excitation

Spatial and temporal variations in the Earth's gravitational field cause perturbations to the motion of satellites as they orbit the Earth. Such orbit perturbations can be determined from ranging measurements, either ranging between ground stations and the satellites as in the case of SLR (see **Section 3.09.3.3.3**) or ranging between satellites as in the case of the GRACE twin satellites (Tapley et al., 2004b). Variations in the Earth's gravitational field are routinely determined from such range measurements, and numerous studies have been conducted that compare the temporal variations in the degree-2 coefficients of the gravitational field to Earth rotation measurements from which motion effects have been modeled and removed.

In a series of studies, J. Chen and colleagues compared degree-2 gravitational field coefficients from SLR and GRACE to Earth rotation measurements and models of surface geophysical fluid mass redistribution (Chen and Wilson, 2003, 2008, 2010, 2012; Chen et al., 2000b, 2004b, 2005, 2012). The availability of three independent estimates of the degree-2 Stokes coefficients (from SLR, GRACE, and EOP via eqns [88] and [89]) and models of surface mass redistribution is essential for verifying and validating the measurements and models and for improving our understanding of the large-scale mass redistribution within and between the atmosphere, oceans, and water stored on land (also see, e.g., Bourda, 2008; Gross et al., 2009; Hancock and Moore, 2007; Jin et al., 2011; Nastula et al., 2007b; Seoane et al., 2009, 2012).

The relatively high spatial resolution of the GRACE gravitational field measurements allows individual Earth rotation excitation processes to be isolated. For example, the individual effects on the Earth's rotation of changes in the mass distribution of the oceans (e.g., Göttl et al., 2012; Jin et al., 2010) and of water stored on land (e.g., Brzezinski et al., 2009; Fernández, 2009; Jin et al., 2010, 2012; Nastula and Salstein, 2012; Seoane et al., 2011) can be studied with GRACE measurements after atmospheric effects are removed from them using a model. The availability of GRACE measurements, in addition to models, of individual Earth rotation excitation processes is leading to major advancements in our knowledge of the relative importance of those processes in causing Earth rotation variations.

Acknowledgments

The work described in this chapter was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Support for this work was provided by the Earth Surface and Interior Focus Area of NASA's Science Mission Directorate.

References

Abarca del Rio R (1999) The influence of global warming in Earth rotation speed. *Annales Geophysicae* 17: 806–811.

Abarca del Rio R and Cazenave A (1994) Interannual variations in the Earth's polar motion for 1963–1991: Comparison with atmospheric angular momentum over 1980–1991. *Geophysical Research Letters* 21(22): 2361–2364.

Abarca del Rio R, Gambis D, and Salstein DA (2000) Interannual signals in length of day and atmospheric angular momentum. *Annales Geophysicae* 18: 347–364.

Alfonsi L and Spada G (1998) Effect of subductions and trends in seismically induced Earth rotational variations. *Journal of Geophysical Research* 103(B4): 7351–7362.

Amit H and Olson P (2004) Helical core flow from geomagnetic secular variation. *Physics of the Earth and Planetary Interiors* 147: 1–25. http://dx.doi.org/10.1016/j.pepi.2004.02. 006.

Amit H and Olson P (2006) Time-average and time-dependent parts of core flow. *Physics of the Earth and Planetary Interiors* 155: 120–139.

Anderson JD and Rosen RD (1983) The latitude-height structure of 40–50 day variations in atmospheric angular momentum. *Journal of Atmospheric Science* 40: 1584–1591.

Aoyama Y (2005) Quasi-14 month wind fluctuation and excitation of the Chandler wobble. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, 24, pp. 135–141. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.

Aoyama Y and Naito I (2000) Wind contributions to the Earth's angular momentum budgets in seasonal variation. *Journal of Geophysical Research* 105(D10): 12417–12431.

Aoyama Y and Naito I (2001) Atmospheric excitation of the Chandler wobble, 1983–1998. *Journal of Geophysical Research* 106(B5): 8941–8954.

Aoyama Y, Naito I, Iwabuchi T, and Yamazaki N (2003) Atmospheric quasi-14 month fluctuation and excitation of the Chandler wobble. *Earth, Planets and Space* 55: e25–e28.

Argus DF and Gross RS (2004) An estimate of motion between the spin axis and the hotspots over the past century. *Geophysical Research Letters* 31: L06614. http://dx.doi.org/10.1029/2004GL019657.

Artz T, Bernhard L, Nothnagel A, Steigenberger P, and Tesmer S (2012) Methodology for the combination of sub-daily Earth rotation from GPS and VLBI observations. *Journal of Geodesy* 86: 221–239. http://dx.doi.org/10.1007/s00190-011-0512-9.

Artz T, Böckmann S, Nothnagel A, and Steigenberger P (2010) Subdaily variations in the Earth's rotation from continuous Very Long Baseline Interferometry campaigns. *Journal of Geophysical Research* 115: B05404. http://dx.doi.org/10.1029/2009JB006834.

Artz T, Tesmer S, and Nothnagel A (2011) Assessment of periodic sub-diurnal Earth rotation variations at tidal frequencies through transformation of VLBI normal equation systems. *Journal of Geodesy* 85: 565–584. http://dx.doi.org/10.1007/s00190-011-0457-z.

Asari S and Lesur V (2011) Radial vorticity constraint in core flow modeling. *Journal of Geophysical Research* 116: B11101. http://dx.doi.org/10.1029/2011JB008267.

Asari S, Shimizu H, and Utada H (2006) Variability of the topographic core–mantle torque calculated from core surface flow models. *Physics of the Earth and Planetary Interiors* 154: 85–111.

Asari S, Shimizu H, and Utada H (2009) Robust and less robust features in the tangential geostrophy core flows. *Geophysical Journal International* 178: 678–692. http://dx.doi.org/10.1111/j.1365-246X.2009.04168.x.

Aubert J and Dumberry M (2011) Steady and fluctuating core rotation in numerical geodynamo models. *Geophysical Journal International* 184: 162–170. http://dx.doi.org/10.1111/j.1365-246X.2010.04842.x.

Baader H-R, Brosche P, and Hovel W (1983) Ocean tides and periodic variations of the Earth's rotation. *Journal of Geophysics* 52: 140–142.

Barnes RTH, Hide R, White AA, and Wilson CA (1983) Atmospheric angular momentum fluctuations, length-of-day changes and polar motion. *Proceedings of the Royal Society of London A* 387: 31–73.

Bellanger E, Blanter EM, Le Mouël J-L, and Shnirman MG (2002a) Estimation of the 13.63-day lunar tide effect on length of day. *Journal of Geophysical Research* 107(B5): 2102.

Bellanger E, Gibert D, and Le Mouël J-L (2002b) A geomagnetic triggering of Chandler wobble phase jumps? *Geophysical Research Letters* 29(7): 1124. http://dx.doi.org/10.1029/2001GL014253.

Bellanger E, Le Mouël J-L, Mandea M, and Labrosse S (2001) Chandler wobble and geomagnetic jerks. *Physics of the Earth and Planetary Interiors* 124: 95–103.

Benjamin D, Wahr J, Ray RD, Egbert GD, and Desai SD (2006) Constraints on mantle anelasticity from geodetic observations, and implications for the *J2* anomaly. *Geophysical Journal International* 165: 3–16. http://dx.doi.org/10.1111/j.1365-246X.2006.02915.x.

Beutler G, Hein GW, Melbourne WG and Seeber G (eds.) (1996) GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications. In: *Proceedings of IAG Symposium no. 115*, p. 351. New York: Springer.

Beutler G, Moore AW, and Mueller II (2009) The international global navigation satellite systems service (IGS): Development and achievements. *Journal of Geodesy* 83(3–4): 297–307. http://dx.doi.org/10.1007/s00190-008-0268-z.

Beutler G, Rothacher M, Schaer S, Springer TA, Kouba J, and Neilan RE (1999) The International GPS Service (IGS): An interdisciplinary service in support of the Earth sciences. *Advances in Space Research* 23(4): 631–653.

Bizouard C and Gambis D (2009) The combined solution C04 for Earth orientation parameters consistent with International Terrestrial Reference Frame 2005. In: Drewes H (ed.) *Geodetic Reference Frames, IAG Symposia*, vol. 134, pp. 265–270. New York: Springer.

Bizouard C, Remus F, Lambert SB, Seoane L, and Gambis D (2011) The Earth's variable Chandler wobble. *Astronomy and Astrophysics* 526: A106. http://dx.doi.org/10.1051/0004-6361/201015894.

Bizouard C and Seoane L (2010) Atmospheric and oceanic forcing of the rapid polar motion. *Journal of Geodesy* 84: 19–30. http://dx.doi.org/10.1007/s00190-009-0341-2.

Blewitt G (1993) Advances in Global Positioning System technology for geodynamics investigations: 1978–1992. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Technology. American Geophysical Union Geodynamics Series*, 25, pp. 195–213. Washington, DC: American Geophysical Union.

Blewitt G (2007) GPS and space-based geodetic methods. In: Herring T (ed.) *Geodesy. Treatise on Geophysics*, vol. 3, pp. 351–390. Amsterdam, Elsevier.

Bloxham J (1990) On the consequences of strong stable stratification at the top of the Earth's outer core. *Geophysical Research Letters* 17(12): 2081–2084.

Bloxham J (1998) Dynamics of angular momentum in the Earth's core. *Annual Review of Earth and Planetary Sciences* 26: 501–517.

Bloxham J and Jackson A (1991) Fluid flow near the surface of the Earth's core. *Reviews of Geophysics* 29(1): 97–120.

Bloxham J and Kuang W (1995) Comment on "The topographic torque on a bounding surface of a rotating gravitating fluid and the excitation by core motions of decadal fluctuations in the Earth's rotation". *Geophysical Research Letters* 22(24): 3561–3562.

Bloxham J, Zatman S, and Dumberry M (2002) The origin of geomagnetic jerks. *Nature* 420: 65–68.

Bock Y and Leppard N (eds.) (1990) Global Positioning System an Overview. In: *Proceedings of IAG Symposium no. 102*, p. 459. New York: Springer.

Böhm S, Brzezinski B, and Schuh H (2012) Complex demodulation in VLBI estimation of high frequency Earth rotation components. *Journal of Geodynamics* 62: 56–68. http://dx.doi.org/10.1016/j.jog.2011.10.002.

Bourda G (2008) Length-of-day and space-geodetic determination of the Earth's variable gravity field. *Journal of Geodesy* 82: 295–305. http://dx.doi.org/10.1007/s00190-007-0180-y.

Braginsky SI (1970) Torsional magnetohydrodynamic vibrations in the Earth's core and variations in day length. *Geomagnetizm I Aeronomiya (English translation)* 10: 1–8.

Brosche P (1982) Oceanic tides and the rotation of the Earth. In: Fricke W and Teleki G (eds.) *Sun and Planetary System*, pp. 179–184. Dordrecht, Holland: Reidel.

Brosche P, Seiler U, Sündermann J, and Wünsch J (1989) Periodic changes in Earth's rotation due to oceanic tides. *Astronomy and Astrophysics* 220: 318–320.

Brosche P and Sündermann J (1985) The Antarctic Circumpolar Current and its influence on the Earth's rotation. *Deutsche Hydrographische Zeitschrift* 38: 1–6.

Brosche P and Wünsch J (1994) On the "rotational angular momentum" of the oceans and the corresponding polar motion. *Astronomische Nachrichten* 315: 181–188.

Brosche P, Wünsch J, Campbell J, and Schuh H (1991) Ocean tide effects in Universal Time detected by VLBI. *Astronomy and Astrophysics* 245: 676–682.

Brosche P, Wünsch J, Frische A, Sündermann J, Maier-Reimer E, and Mikolajewicz U (1990) The seasonal variation of the angular momentum of the oceans. *Naturwissenschaften* 77: 185–186.

Brosche P, Wünsch J, Maier-Reimer E, Segschneider J, and Sündermann J (1997) The axial angular momentum of the general circulation of the oceans. *Astronomische Nachrichten* 318: 193–199.

Brzezinski A (1987) Statistical investigations on atmospheric angular momentum functions and on their effects on polar motion. *Manuscripta Geodaetica* 12: 268–281.

Brzezinski A (1992) Polar motion excitation by variations of the effective angular momentum function: Considerations concerning deconvolution problem. *Manuscripta Geodaetica* 17: 3–20.

Brzezinski A, Bizouard Ch, and Petrov S (2002a) Influence of the atmosphere on Earth rotation: What new can be learned from the recent atmospheric angular momentum estimates? *Surveys in Geophysics* 23: 33–69.

Brzezinski A and Capitaine N (1993) The use of the precise observations of the celestial ephemeris pole in the analysis of geophysical excitation of Earth rotation. *Journal of Geophysical Research* 98(B4): 6667–6675.

Brzezinski A, Dobslaw H, Dill R, and Thomas M (2012) Geophysical excitation of the Chandler wobble revisited. In: Kenyon S, Pacino MC, and Marti UJ (eds.) *Geodesy for Planet Earth, IAG Symposia*, vol. 136, pp. 499–505. New York: Springer.

Brzezinski A and Nastula J (2002) Oceanic excitation of the Chandler wobble. *Advances in Space Research* 30: 195–200.

Brzezinski A, Nastula J, and Kolaczek B (2009) Seasonal excitation of polar motion estimated from recent geophysical models and observations. *Journal of Geodynamics* 48: 235–240. http://dx.doi.org/10.1016/j.jog.2009.09.021.

Brzezinski A, Nastula J, Kolaczek B, and Ponte RM (2005) Oceanic excitation of polar motion from intraseasonal to decadal periods. In: Sansò F (ed.) *A Window on the Future of Geodesy, IAG Symposia*, vol. 128, pp. 591–596. Berlin: Springer.

Brzezinski A, Nastula J, and Ponte RM (2002b) Oceanic excitation of the Chandler wobble using a 50-year time series of ocean angular momentum. In: Adám J and Schwarz K-P (eds.) *Vistas for Geodesy in the New Millennium, IAG Symposia*, vol. 125, pp. 434–439. Berlin: Springer.

Brzezinski A, Ponte RM, and Ali AH (2004) Nontidal oceanic excitation of nutation and diurnal/semidiurnal polar motion revisited. *Journal of Geophysical Research* 109: B11407. http://dx.doi.org/10.1029/2004JB003054.

Buffett BA (1996a) Gravitational oscillations in the length of day. *Geophysical Research Letters* 23(17): 2279–2282.

Buffett BA (1996b) A mechanism for decade fluctuations in the length of day. *Geophysical Research Letters* 23(25): 3803–3806.

Buffett BA (1998) Free oscillations in the length of day: Inferences on physical properties near the core–mantle boundary. In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region. American Geophysical*

Union Geodynamics Series, vol. 28, pp. 153–165. Washington, DC, American Geophysical Union.

Buffett BA (2007) Core–mantle interactions. In: Olson P (ed.) *Treatise on Geophysics*, vol. 8, pp. 345–358. Amsterdam, Elsevier.

Buffett BA and Creager KC (1999) A comparison of geodetic and seismic estimates of inner-core rotation. *Geophysical Research Letters* 26(10): 1509–1512.

Buffett BA and Mound JE (2005) A Green's function for the excitation of torsional oscillations in the Earth's core. *Journal of Geophysical Research* 110: B08104. http://dx.doi.org/10.1029/2004JB003495.

Buffett BA, Mound J, and Jackson A (2009) Inversion of torsional oscillations for the structure and dynamics of Earth's core. *Geophysical Journal International* 177: 878–890. http://dx.doi.org/10.1111/j.1365-246X.2009.04129.x.

Bullard EC, Freeman C, Gellman H, and Nixon J (1950) The westward drift of the Earth's magnetic field. *Philosophical Transactions of the Royal Society of London A* 243: 61–92.

Cambiotti G, Ricard Y, and Sabadini R (2010) Ice age True Polar Wander in a compressible and non-hydrostatic Earth. *Geophysical Journal International* 183: 1248–1264. http://dx.doi.org/10.1111/j.1365-246X.2010.04791.x.

Cambiotti G, Ricard Y, and Sabadini R (2011) New insights into mantle convection true polar wander and rotational bulge readjustment. *Earth and Planetary Science Letters* 310: 538–543. http://dx.doi.org/10.1016/j.epsl.2011.08.009.

Canet E, Fournier A, and Jault D (2009) Forward and adjoint quasi-geostrophic models of the geomagnetic secular variation. *Journal of Geophysical Research* 114: B11101. http://dx.doi.org/10.1029/2008JB006189.

Capitaine N (2000) Definition of the celestial ephemeris pole and the celestial ephemeris origin. In: Johnston KJ, McCarthy DD, Luzum BJ, and Kaplan GH (eds.) *Towards Models and Constants for Sub-Microarcsecond Astrometry, Proc. IAU Colloquium 180, US Naval Obs*, pp. 153–163. Washington, DC: US Naval Observatory.

Celaya MA, Wahr JM, and Bryan FO (1999) Climate-driven polar motion. *Journal of Geophysical Research* 104(B6): 12813–12829.

Chan N-H, Mitrovica JX, Matsuyama I, Creveling JR, and Stanley S (2011a) The rotational stability of a convecting Earth: Assessing inferences of rapid TPW in the Late Cretaceous. *Geophysical Journal International* 187: 1319–1333. http://dx.doi.org/10.1111/j.1365-246X.2011.05245.x.

Chan N-H, Mitrovica JX, Matsuyama I, et al. (2011b) The rotational stability of a convecting Earth: The Earth's figure and TPW over the last 100 Myr. *Geophysical Journal International* 187: 773–782. http://dx.doi.org/10.1111/j.1365-246X.2011.05174.x.

Chandler SC (1891) On the variation of latitude, II. *The Astronomical Journal* 11: 65–70.

Chandler SC (1892) On the variation of latitude, VII. *The Astronomical Journal* 12: 97–101.

Chao BF (1983) Autoregressive harmonic analysis of the Earth's polar motion using homogeneous International Latitude Service data. *Journal of Geophysical Research* 88(B12): 10299–10307.

Chao BF (1984) Interannual length-of-day variations with relation to the Southern Oscillation/El Niño. *Geophysical Research Letters* 11(5): 541–544.

Chao BF (1988) Correlation of interannual length-of-day variation with El Niño/Southern Oscillation, 1972–1986. *Journal of Geophysical Research* 93(B7): 7709–7715.

Chao BF (1989) Length-of-day variations caused by El Niño/Southern Oscillation and Quasi-Biennial Oscillation. *Science* 243: 923–925.

Chao BF (1993) Excitation of Earth's polar motion by atmospheric angular momentum variations, 1980–1990. *Geophysical Research Letters* 20(2): 253–256.

Chao BF (1994) Zonal tidal signals in the Earth's polar motion. *Eos, Transactions of the American Geophysical Union* 75(44): 158.

Chao BF and Au AY (1991) Atmospheric excitation of the Earth's annual wobble: 1980–1988. *Journal of Geophysical Research* 96(B4): 6577–6582.

Chao BF and Gross RS (1987) Changes in the Earth's rotation and low-degree gravitational field introduced by earthquakes. *Geophysical Journal of the Royal Astronomical Society* 91: 569–596.

Chao BF, Gross RS, and Han Y (1996a) Seismic excitation of the polar motion, 1977–1993. *Pure and Applied Geophysics* 146(3/4): 407–419.

Chao BF, Merriam JB, and Tamura Y (1995a) Geophysical analysis of zonal tidal signals in length of day. *Geophysical Journal International* 122: 765–775.

Chao BF and O'Connor WP (1988) Global surface-water-induced seasonal variations in the Earth's rotation and gravitational field. *Geophysical Journal of the Royal Astronomical Society* 94: 263–270.

Chao BF, O'Connor WP, Chang ATC, Hall DK, and Foster JL (1987) Snow load effect on the Earth's rotation and gravitational field, 1979–1985. *Journal of Geophysical Research* 92(B9): 9415–9422.

Chao BF and Ray RD (1997) Oceanic tidal angular momentum and Earth's rotation variations. *Progress in Oceanography* 40: 399–421.

Chao BF, Ray RD, and Egbert GD (1995b) Diurnal/semidiurnal oceanic tidal angular momentum: Topex/Poseidon models in comparison with Earth's rotation rate. *Geophysical Research Letters* 22(15): 1993–1996.

Chao BF, Ray RD, Gipson JM, Egbert GD, and Ma C (1996b) Diurnal/semidiurnal polar motion excited by oceanic tidal angular momentum. *Journal of Geophysical Research* 101(B9): 20151–20163.

Chao BF and Salstein DA (2005) Mass, momentum, and geodynamics. In: Lau WKM and Waliser DE (eds.) *Intraseasonal Variability in the Atmosphere—Ocean Climate System*, pp. 247–269. Berlin: Springer.

Chao BF and Yan H (2010) Relation between length-of-day variation and angular momentum of geophysical fluids. *Journal of Geophysical Research* 115: B10417. http://dx.doi.org/10.1029/2009JB007024.

Chao BF and Zhou Y-H (1999) Meteorological excitation of interannual polar motion by the North Atlantic Oscillation. *Journal of Geodynamics* 27: 61–73.

Chapman S and Lindzen RS (1970) *Atmospheric Tides*. Dordrecht, The Netherlands: D. Reidel.

Chapront J, Chapront-Touzé M, and Francou G (2002) A new determination of lunar orbital parameters, precession constant and tidal acceleration from LLR measurements. *Astronomy and Astrophysics* 387: 700–709.

Chen J (2005) Global mass balance and the length-of-day variations. *Journal of Geophysical Research* 110: B08404. http://dx.doi.org/10.1029/2004JB003474.

Chen W and Shen W (2010) New estimates of the inertia tensor and rotation of the triaxial nonrigid Earth. *Journal of Geophysical Research* 115: B12419. http://dx.doi.org/10.1029/2009JB 007094.

Chen JL and Wilson CR (2003) Low degree gravitational changes from Earth rotation and geophysical models. *Geophysical Research Letters* 30(24): 2257. http://dx.doi.org/10.1029/2003GL018688.

Chen JL and Wilson CR (2005) Hydrological excitations of polar motion, 1993–2002. *Geophysical Journal International* 160: 833–839.

Chen JL and Wilson CR (2008) Low degree gravity changes from GRACE, Earth rotation, geophysical models, and satellite laser ranging. *Journal of Geophysical Research* 113: B06402. http://dx.doi.org/10.1029/2007JB005397.

Chen JL and Wilson CR (2010) Assessment of degree-2 zonal gravitational changes from GRACE, Earth rotation, climate models, and satellite laser ranging. In: Mertikas SP (ed.) *Gravity, Geoid, and Earth Observation, IAG Symposia*, vol. 135, pp. 669–676. Berlin: Springer.

Chen JL and Wilson CR (2012) Multi-sensor monitoring of low-degree gravitational changes. In: Sneeuw N, Novák P, Crespi M, and Sansò F (eds.) *VII Hotine-Marussi Symposium on Mathematical Geodesy, IAG Symposia*, vol. 137, pp. 293–300. Berlin: Springer.

Chen JL, Wilson CR, Chao BF, Shum CK, and Tapley BD (2000a) Hydrological and oceanic excitations to polar motion and length-of-day variation. *Geophysical Journal International* 141: 149–156.

Chen JL, Wilson CR, Eanes RJ, and Tapley BD (2000b) A new assessment of long-wavelength gravitational variations. *Journal of Geophysical Research* 105(B7): 16271–16277.

Chen JL, Wilson CR, Hu X-G, Zhou Y-H, and Tapley BD (2004a) Oceanic effects on polar motion determined from an ocean model and satellite altimetry: 1993–2001. *Journal of Geophysical Research* 109: B02411. http://dx.doi.org/10.1029/2003JB002664.

Chen JL, Wilson CR, and Tapley BD (2005) Interannual variability of low-degree gravitational change, 1980–2002. *Journal of Geodesy* 78: 535–543. http://dx.doi.org/10.1007/s00190-004-0417-y.

Chen JL, Wilson CR, Tapley BD, and Ries JC (2004b) Low degree gravitational changes from GRACE: Validation and interpretation. *Geophysical Research Letters* 31: L22607. http://dx.doi.org/10.1029/2004GL021670.

Chen JL, Wilson CR, and Zhou YH (2012) Seasonal excitation of polar motion. *Journal of Geodynamics* 62: 8–15. http://dx.doi.org/10.1016/j.jog.2011.12.002.

Cheng MK, Eanes RJ, and Tapley BD (1992) Tidal deceleration of the Moon's mean motion. *Geophysical Journal International* 108: 401–409.

Cheng MK, Shum CK, Eanes RJ, Schutz BE, and Tapley BD (1990) Long-period perturbations in Starlette orbit and tide solution. *Journal of Geophysical Research* 95(B6): 8723–8736.

Christodoulidis DC, Smith DE, Williamson RG, and Klosko SM (1988) Observed tidal braking in the Earth/Moon/Sun system. *Journal of Geophysical Research* 93(B6): 6216–6236.

Combrinck L (2010) Satellite laser ranging. In: Xu G (ed.) *Sciences of Geodesy I: Advances and Future Directions*, pp. 301–338. New York: Springer.

Dahlen FA (1976) The passive influence of the oceans upon the rotation of the Earth. *Geophysical Journal of the Royal Astronomical Society* 46: 363–406.

Dai A and Wang J (1999) Diurnal and semidiurnal tides in global surface pressure fields. *Journal of Atmospheric Science* 56: 3874–3891.

Dalmau W (1997) Critical remarks on the use of medieval eclipse records for the determination of long-term changes in the Earth's rotation. *Surveys in Geophysics* 18: 213–223.
Davis RG and Whaler KA (1996) Determination of a steady velocity in a rotating frame of reference at the surface of the Earth's core. *Geophysical Journal International* 126: 92–100.
de Viron O, Bizouard C, Salstein D, and Dehant V (1999) Atmospheric torque on the Earth and comparison with atmospheric angular momentum variations. *Journal of Geophysical Research* 104(B3): 4861–4875.
de Viron O, Boy J-P, and Goosse H (2004a) Geodetic effects of the ocean response to atmospheric forcing in an ocean general circulation model. *Journal of Geophysical Research* 109: B03411. http://dx.doi.org/10.1029/2003JB002837.
de Viron O and Dehant V (1999) Earth's rotation and high frequency equatorial angular momentum budget of the atmosphere. *Surveys in Geophysics* 20: 441–462.
de Viron O and Dehant V (2003) Tests on the validity of atmospheric torques on Earth computed from atmospheric model outputs. *Journal of Geophysical Research* 108(B2): 2068. http://dx.doi.org/10.1029/2001JB001196.
de Viron O, Dehant V, Goosse H, Crucifix M, Participating CMIP, and Groups Modeling (2002) Effect of global warming on the length-of-day. *Geophysical Research Letters* 29(7): 1146. http://dx.doi.org/10.1029/2001GL013672.
de Viron O, Marcus SL, and Dickey JO (2001a) Diurnal angular momentum budget of the atmosphere and its consequences for Earth's nutation. *Journal of Geophysical Research* 106(B11): 26747–26759.
de Viron O, Marcus SL, and Dickey JO (2001b) Atmospheric torques during the winter of 1989: Impact of ENSO and NAO positive phases. *Geophysical Research Letters* 28(10): 1985–1988.
de Viron O, Salstein D, Bizouard Ch, and Fernandez L (2004b) Low-frequency excitation of length of day and polar motion by the atmosphere. *Journal of Geophysical Research* 109: B03408. http://dx.doi.org/10.1029/2003JB002817.
de Viron O, Schwarzbaum G, Lott F, and Dehant V (2005) Diurnal and subdiurnal effects of the atmosphere on the Earth rotation and geocenter motion. *Journal of Geophysical Research* 110: B11404. http://dx.doi.org/10.1029/2005JB003761.
Defraigne P and Smits I (1999) Length of day variations due to zonal tides for an inelastic Earth in non-hydrostatic equilibrium. *Geophysical Journal International* 139: 563–572.
Desai SD and Wahr JM (1995) Empirical ocean tide models estimated from Topex/Poseidon altimetry. *Journal of Geophysical Research* 100(C12): 25205–25228.
Desai SD and Wahr JM (1999) Monthly and fortnightly tidal variations of the Earth's rotation rate predicted by a Topex/Poseidon empirical ocean tide model. *Geophysical Research Letters* 26(8): 1035–1038.
di Leonardo SM and Dickman SR (2004) Isolation of atmospheric effects on rapid polar motion through Wiener filtering. *Geophysical Journal International* 159: 863–873.
Dick WR and Richter B (2004) The International Earth Rotation and Reference Systems Service (IERS). In: Heck A (ed.) *Organizations and Strategies in Astronomy,* vol. 5, pp. 159–168. Dordrecht: Kluwer.
Dickey JO, Bender PL, Faller JE, et al. (1994a) Lunar laser ranging: A continuing legacy of the Apollo program. *Science* 265: 482–490.
Dickey JO and de Viron O (2009) Leading modes of torsional oscillations within the Earth's core. *Geophysical Research Letters* 36: L15302. http://dx.doi.org/10.1029/GL038386.
Dickey JO, Gegout P, and Marcus SL (1999) Earth-atmosphere angular momentum exchange and ENSO: The rotational signature of the 1997–98 event. *Geophysical Research Letters* 26(16): 2477–2480.
Dickey JO, Ghil M, and Marcus SL (1991) Extratropical aspects of the 40–50 day oscillation in length-of-day and atmospheric angular momentum. *Journal of Geophysical Research* 96(D12): 22643–22658.
Dickey JO, Marcus SL, and Chin TM (2007) Thermal wind forcing and atmospheric angular momentum: Origin of the Earth's delayed response to ENSO. *Geophysical Research Letters* 34: L17803. http://dx.doi.org/10.1029/2007GL030846.
Dickey JO, Marcus SL, and de Viron O (2010) Closure in the Earth's angular momentum budget observed from subseasonal periods down to four days: No core effects needed. *Geophysical Research Letters* 37: L03307. http://dx.doi.org/10.1029/2009GL041118.
Dickey JO, Marcus SL, Eubanks TM, and Hide R (1993a) Climate studies via space geodesy: Relationships between ENSO and interannual length-of-day variations. In: McBean GA and Hantel M (eds.) *Interactions Between Global Climate Subsystems: The Legacy of Hann. American Geophysical Union Geophysical Monograph Series,* 75, pp. 141–155. American Geophysical Union.
Dickey JO, Marcus SL, and Hide R (1992a) Global propagation of interannual fluctuations in atmospheric angular momentum. *Nature* 357: 482–488.
Dickey JO, Marcus SL, Hide R, Eubanks TM, and Boggs DH (1994b) Angular momentum exchange among the solid Earth, atmosphere and oceans: A case study of the 1982–83 El Niño event. *Journal of Geophysical Research* 99(B12): 23921–23937.
Dickey JO, Marcus SL, Johns CM, Hide R, and Thompson SR (1993b) The oceanic contribution to the Earth's seasonal angular momentum budget. *Geophysical Research Letters* 20: 2953–2956.
Dickey JO, Marcus SL, Steppe JA, and Hide R (1992b) The Earth's angular momentum budget on subseasonal time scales. *Science* 255: 321–324.
Dickey JO, Newhall XX, and Williams JG (1985) Earth orientation from lunar laser ranging and an error analysis of polar motion services. *Journal of Geophysical Research* 90: 9353–9362.
Dickman SR (1979) Continental drift and true polar wandering. *Geophysical Journal of the Royal Astronomical Society* 57: 41–50.
Dickman SR (1981) Investigation of controversial polar motion features using homogeneous International Latitude Service data. *Journal of Geophysical Research* 86(B6): 4904–4912.
Dickman SR (1993) Dynamic ocean-tide effects on Earth's rotation. *Geophysical Journal International* 112: 448–470.
Dickman SR (1994) Ocean tidal effects on Earth's rotation and on the lunar orbit. In: Schutz BE, Anderson A, Froidevaux C, and Parke M (eds.) *Gravimetry and Space Techniques Applied to Geodynamics and Ocean Dynamics. American Geophysical Union Geophysical Monograph Series*, vol. 82, pp. 87–94. Washington, DC, American Geophysical Union.
Dickman SR (2003) Evaluation of "effective angular momentum function" formulations with respect to core–mantle coupling. *Journal of Geophysical Research* 108(B3): 2150. http://dx.doi.org/10.1029/2001JB001603.
Dickman SR (2005) Rotationally consistent Love numbers. *Geophysical Journal International* 161: 31–40.
Dickman SR and Gross RS (2010) Rotational evaluation of a long-period spherical harmonic ocean tide model. *Journal of Geodesy* 84: 457–464. http://dx.doi.org/10.1007/s00190-010-0383-5.
Dickman SR and Nam YS (1995) Revised predictions of long-period ocean tidal effects on Earth's rotation rate. *Journal of Geophysical Research* 100(B5): 8233–8243.
Djurovic D and Pâquet P (1996) The common oscillations of solar activity, the geomagnetic field, and the Earth's rotation. *Solar Physics* 167: 427–439.
Dobslaw H, Dill R, Grötzsch A, Brzezinski A, and Thomas M (2010) Seasonal polar motion excitation from numerical models of atmosphere, ocean, and continental hydrosphere. *Journal of Geophysical Research* 115: B10406. http://dx.doi.org/10.1029/2009JB007127.
Dong D and Herring T (1990) Observed variations of UT1 and polar motion in the diurnal and semidiurnal bands. *Eos, Transactions of the American Geophysical Union* 71: 482.
Dow JM, Neilan RE, and Gendt G (2005) The International GPS Service: Celebrating the 10th anniversary and looking to the next decade. *Advances in Space Research* 36(3): 320–326. http://dx.doi.org/10.1016/j.asr.2005.05.125.
Dow JM, Neilan RE, and Rizos C (2009) The International GNSS Service in a changing landscape of Global Navigation Satellite Systems. *Journal of Geodesy* 83(3–4): 191–198. http://dx.doi.org/10.1007/s00190-008-0300-3.
Dumberry M (2007) Geodynamic constraints on the steady and time-dependent inner core axial rotation. *Geophysical Journal International* 170: 886–895. http://dx.doi.org/10.1111/j.1365-246X.2007.03484.x.
Dumberry M (2008) Gravitational torque on the inner core and decadal polar motion. *Geophysical Journal International* 172: 903–920. http://dx.doi.org/10.1111/j.1365-246X.2007.03653.x.
Dumberry M (2010) Gravity variations induced by core flows. *Geophysical Journal International* 180: 635–650. http://dx.doi.org/10.1111/j.1365-246X.2009.04437.x.
Dumberry M and Bloxham J (2002) Inner core tilt and polar motion. *Geophysical Journal International* 151: 377–392.
Dumberry M and Bloxham J (2004) Variations in the Earth's gravity field caused by torsional oscillations in the core. *Geophysical Journal International* 159: 417–434.
Dumberry M and Bloxham J (2006) Azimuthal flows in the Earth's core and changes in length of day at millennial timescales. *Geophysical Journal International* 165: 32–46. http://dx.doi.org/10.1111/j.1365-246X.2006.02903.x.
Dumberry M and Mound JE (2008) Constraints on core–mantle electromagnetic coupling from torsional oscillation normal modes. *Journal of Geophysical Research* 113: B03102. http://dx.doi.org/10.1029/2007JB005135.
Dumberry M and Mound J (2010) Inner core–mantle gravitational locking and the super-rotation of the inner core. *Geophysical Journal International* 181: 806–817. http://dx.doi.org/10.1111/j.1365-246X.2010.04563.x.
Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.
Eubanks TM (1993) Variations in the orientation of the Earth. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Earth*

Dynamics. *American Geophysical Union Geodynamics Series*, 24, pp. 1–54. Washington, DC: American Geophysical Union.

Eubanks TM, Steppe JA, and Dickey JO (1986) The El Niño, the Southern Oscillation and the Earth's rotation. In: Cazenave A (ed.) *Earth Rotation: Solved and Unsolved Problems*, pp. 163–186. Hingham, MA: D. Reidel.

Eubanks TM, Steppe JA, Dickey JO, and Callahan PS (1985) A spectral analysis of the Earth's angular momentum budget. *Journal of Geophysical Research* 90(B7): 5385–5404.

Eubanks TM, Steppe JA, Dickey JO, Rosen RD, and Salstein DA (1988) Causes of rapid motions of the Earth's pole. *Nature* 334: 115–119.

Euler L (1765) *Theoria Motus Corporum Solidorum*. Rostock, Germany: Litteris et impensis A.F. Röse.

Exertier P, Bonnefond P, Deleflie F, et al. (2006) Contribution of laser ranging to Earth's sciences. *Comptes Rendus Geoscience* 338: 958–967. http://dx.doi.org/10.1016/j.crte.2006.09.019.

Fang M, Hager BH, and Herring TA (1996) Surface deformation caused by pressure changes in the fluid core. *Geophysical Research Letters* 23(12): 1493–1496.

Feissel M and Gambis D (1980) La mise en evidence de variations rapides de la durée du jour. *Comptes Rendus De L Academie Des Sciences Paris, Series B* 291: 271–273.

Feissel M and Gambis D (1993) The International Earth Rotation Service: Current results for research on Earth rotation and reference frames. *Advances in Space Research* 13(11): 143–150.

Feissel M and Gavoret J (1990) ENSO-related signals in Earth rotation, 1962–87. In: McCarthy DD and Carter WE (eds.) *Variations in Earth Rotation*. *American Geophysical Union Geophysical Monograph Series*, vol. 59, pp. 133–137. Washington, DC, American Geophysical Union.

Feissel M and Nitschelm C (1985) Time dependent aspects of the atmospheric driven fluctuations in the duration of the day. *Annales Geophysicae* 3: 180–186.

Feldstein SB (2006) Dynamical processes of equatorial atmospheric angular momentum. *Journal of the Atmospheric Sciences* 63: 565–581.

Feldstein SB (2008) The dynamics of atmospherically driven intraseasonal polar motion. *Journal of the Atmospheric Sciences* 65: 2290–2307. http://dx.doi.org/10.1175/2007JAS2640.1.

Fernández LI (2009) Analysis of geophysical variations of the *C20* coefficient of the geopotential. In: Sideris MG (ed.) *Observing Our Changing Earth, IAG Symposia,* vol. 133, pp. 493–500. Berlin: Springer.

Fernández LI, Schuh H, Schmidt M, and Seitz F (2007) Effects of inter-annual water storage variations on polar motion. *Geophysical Journal International* 169: 12–18. http://dx.doi.org/10.1111/j.1365-246X.2006.03304.x.

Finlay CC, Dumberry M, Chulliat A, and Pais MA (2010) Short timescale core dynamics: Theory and observations. *Space Science Reviews* 155: 177–218. http://dx.doi.org/10.1007/s11214-010-9691-6.

Fliegel HF and Hawkins TP (1967) Analysis of variations in the rotation of the Earth. *The Astronomical Journal* 72(4): 544–550.

Freedman AP, Ibañez-Meier R, Herring TA, Lichten SM, and Dickey JO (1994a) Subdaily Earth rotation during the Epoch '92 campaign. *Geophysical Research Letters* 21(9): 769–772.

Freedman AP, Steppe JA, Dickey JO, Eubanks TM, and Sung L-Y (1994b) The short-term prediction of universal time and length of day using atmospheric angular momentum. *Journal of Geophysical Research* 99(B4): 6981–6996.

Frische A and Sündermann J (1990) The seasonal angular momentum of the thermohaline ocean circulation. In: Brosche P and Sündermann J (eds.) *Earth's Rotation from Eons to Days*, pp. 108–126. New York: Springer.

Furuya M and Chao BF (1996) Estimation of period and *Q* of the Chandler wobble. *Geophysical Journal International* 127: 693–702.

Furuya M and Hamano Y (1998) Effect of the Pacific Ocean on the Earth's seasonal wobble inferred from National Center for Environmental Prediction ocean analysis data. *Journal of Geophysical Research* 103(B5): 10131–10140.

Furuya M, Hamano Y, and Naito I (1996) Quasi-periodic wind signal as a possible excitation of Chandler wobble. *Journal of Geophysical Research* 101(B11): 25537–25546.

Furuya M, Hamano Y, and Naito I (1997) Importance of wind for the excitation of Chandler wobble as inferred from wobble domain analysis. *Journal of Physics of the Earth* 45(3): 177–188.

Gambis D (1992) Wavelet transform analysis of the length of the day and the El-Niño/Southern Oscillation variations at intraseasonal and interannual time scales. *Annales Geophysicae* 10: 429–437.

Gambis D (2004) Monitoring Earth orientation using space-geodetic techniques: State-of-the-art and prospective. *Journal of Geodesy* 78: 295–303.

Gambis D (2006) DORIS and the determination of the Earth's polar motion. *Journal of Geodesy* 80: 649–656. http://dx.doi.org/10.1007/s00190-006-0043-y.

Gambis D and Luzum B (2011) Earth rotation monitoring, UT1 determination and prediction. *Metrologia* 48: S165–S170. http://dx.doi.org/10.1088/0026-1394/48/4/S06.

Gasperini P, Sabadini R, and Yuen DA (1986) Excitation of the Earth's rotational axis by recent glacial discharges. *Geophysical Research Letters* 13(6): 533–536.

Gibert D, Holschneider M, and Le Mouël J-L (1998) Wavelet analysis of the Chandler wobble. *Journal of Geophysical Research* 103(B11): 27069–27089.

Gibert D and Le Mouël J-L (2008) Inversion of polar motion data: Chandler wobble, phase jumps, and geomagnetic jerks. *Journal of Geophysical Research* 113: B10405. http://dx.doi.org/10.1029/2008JB005700.

Gilbert F and Dziewonski AM (1975) An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra. *Philosophical Transactions of the Royal Society of London A* 278: 187–269.

Gillet N, Jault D, Canet E, and Fournier A (2010) Fast torsional waves and strong magnetic field within the Earth's core. *Nature* 465: 74–77. http://dx.doi.org/10.1038/nature09010.

Gillet N, Pais MA, and Jault D (2009) Ensemble inversion of time-dependent core flow models. *Geochemistry, Geophysics, Geosystems* 10(6): Q06004. http://dx.doi.org/10.1029/2008GC002290.

Gipson JM (1996) Very long baseline interferometry determination of neglected tidal terms in high-frequency Earth orientation variation. *Journal of Geophysical Research* 101(B12): 28051–28064.

Gire C and Le Mouël J-L (1986) Flow in the fluid core and Earth's rotation. In: Cazenave A (ed.) *Earth Rotation: Solved and Unsolved Problems*, pp. 241–258. Dordrecht, Holland: D. Reidel.

Goldstein H (1950) *Classical Mechanics*. Reading, MA: Addison-Wesley.

Göttl F, Schmidt M, Heinkelmann R, Savcenko R, and Bouman J (2012) Combination of gravimetric and altimetric space observations for estimating oceanic polar motion excitations. *Journal of Geophysical Research* 117: C10022. http://dx.doi.org/10.1029/2012JC007915.

Göttl F and Seitz F (2009) Contribution of non-tidal oceanic mass variations to polar motion determined from space geodesy and ocean data. In: Sideris MG (ed.) *Observing Our Changing Earth, IAG Symposia,* vol. 133, pp. 439–445. Berlin: Springer.

Greff-Lefftz M (2004) Upwelling mantle plumes, superswells, and true polar wander. *Geophysical Journal International* 159: 1125–1137.

Greff-Lefftz M (2011) Length of day variations due to mantle dynamics at geological timescale. *Geophysical Journal International* 187: 595–612. http://dx.doi.org/10.1111/j.1365-246X.2011.05169.x.

Greff-Lefftz M and Legros H (1995) Core–mantle coupling and polar motion. *Physics of the Earth and Planetary Interiors* 91: 273–283.

Greff-Lefftz M, Pais MA, and Le Mouël J-L (2004) Surface gravitational field and topography changes induced by the Earth's fluid core motions. *Journal of Geodesy* 78: 386–392.

Greiner-Mai H and Barthelmes F (2001) Relative wobble of the Earth's inner core derived from polar motion and associated gravity variations. *Geophysical Journal International* 144: 27–36.

Greiner-Mai H, Jochmann H, and Barthelmes F (2000) Influence of possible inner-core motions on the polar motion and the gravity field. *Physics of the Earth and Planetary Interiors* 117: 81–93.

Greiner-Mai H, Jochmann H, Barthelmes F, and Ballani L (2003) Possible influences of core processes on the Earth's rotation and the gravity field. *Journal of Geodynamics* 36: 343–358.

Grejner-Brzezinska DA and Goad CC (1996) Subdaily Earth rotation determined from GPS. *Geophysical Research Letters* 23(19): 2701–2704.

Gross RS (1986) The influence of earthquakes on the Chandler wobble during 1977–1983. *Geophysical Journal of the Royal Astronomical Society* 85: 161–177.

Gross RS (1992) Correspondence between theory and observations of polar motion. *Geophysical Journal International* 109: 162–170.

Gross RS (1993) The effect of ocean tides on the Earth's rotation as predicted by the results of an ocean tide model. *Geophysical Research Letters* 20(4): 293–296.

Gross RS (2000a) Combinations of Earth orientation measurements: SPACE97, COMB97, and POLE97. *Journal of Geodesy* 73: 627–637.

Gross RS (2000b) The excitation of the Chandler wobble. *Geophysical Research Letters* 27(15): 2329–2332.

Gross RS (2001) A combined length-of-day series spanning 1832–1997: LUNAR97. *Physics of the Earth and Planetary Interiors* 123: 65–76.

Gross RS (2004) Angular momentum in the Earth system. In: Sanso F (ed.) *V Hotine-Marussi Symposium on Mathematical Geodesy, IAG Symposia,* vol. 127, pp. 274–284. New York: Springer.

Gross RS (2005a) Oceanic excitation of polar motion: A review. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24,

pp. 89–102. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.
Gross RS (2005b) The observed period and Q of the Chandler wobble. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 31–37. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.
Gross RS (2009a) Ocean tidal effects on Earth rotation. *Journal of Geodynamics* 48: 219–225. http://dx.doi.org/10.1016/j.jog.2009.09.016.
Gross RS (2009b) An improved empirical model for the effect of long-period ocean tides on polar motion. *Journal of Geodesy* 83: 635–644. http://dx.doi.org/10.1007/s00190-008-0277-y.
Gross RS and Chao BF (2006) The rotational and gravitational signature of the December 26, 2004 Sumatra earthquake. *Surveys in Geophysics* 27: 615–632. http://dx.doi.org/10.1007/s10712-006-9008-1.
Gross RS, Chao BF, and Desai S (1997) Effect of long-period ocean tides on the Earth's polar motion. *Progress in Oceanography* 40: 385–397.
Gross RS, Eubanks TM, Steppe JA, Freedman AP, Dickey JO, and Runge TF (1998) A Kalman-filter-based approach to combining independent Earth-orientation series. *Journal of Geodesy* 72: 215–235.
Gross RS, Fukumori I, and Menemenlis D (2003) Atmospheric and oceanic excitation of the Earth's wobbles during 1980–2000. *Journal of Geophysical Research* 108(B8): 2370. http://dx.doi.org/10.1029/2002JB002143.
Gross RS, Fukumori I, and Menemenlis D (2005) Atmospheric and oceanic excitation of decadal-scale Earth orientation variations. *Journal of Geophysical Research* 110: B09405. http://dx.doi.org/10.1029/2004JB003565.
Gross RS, Fukumori I, Menemenlis D, and Gegout P (2004) Atmospheric and oceanic excitation of length-of-day variations during 1980–2000. *Journal of Geophysical Research* 109: B01406. http://dx.doi.org/10.1029/2003JB002432.
Gross RS, Hamdan KH, and Boggs DH (1996a) Evidence for excitation of polar motion by fortnightly ocean tides. *Geophysical Research Letters* 23(14): 1809–1812.
Gross RS, Lavallée DA, Blewitt G, and Clarke PJ (2009) Consistency of Earth rotation, gravity, and shape measurements. In: Sideris MG (ed.) *Observing Our Changing Earth, IAG Symposia,* vol. 133, pp. 463–471. Berlin: Springer.
Gross RS and Lindqwister UJ (1992) Atmospheric excitation of polar motion during the GIG '91 measurement campaign. *Geophysical Research Letters* 19(9): 849–852.
Gross RS, Marcus SL, and Dickey JO (2002) Modulation of the seasonal cycle in length-of-day and atmospheric angular momentum. In: Adám J and Schwarz K-P (eds.) *Vistas for Geodesy in the New Millennium, IAG Symposia,* vol. 125, pp. 457–462. Berlin: Springer.
Gross RS, Marcus SL, Eubanks TM, Dickey JO, and Keppenne CL (1996b) Detection of an ENSO signal in seasonal length-of-day variations. *Geophysical Research Letters* 23(23): 3373–3376.
Gross RS and Vondrák J (1999) Astrometric and space-geodetic observations of polar wander. *Geophysical Research Letters* 26(14): 2085–2088.
Groten E (2004) Fundamental parameters and current (2004) best estimates of the parameters of common relevance to astronomy, geodesy, and geodynamics. *Journal of Geodesy* 77: 724–731.
Gubbins D (1982) Finding core motions from magnetic observations. *Philosophical Transactions of the Royal Society of London A* 306: 247–254.
Guinot B (1970) Short-period terms in Universal Time. *Astronomy and Astrophysics* 8: 26–28.
Gurtner W, Noomen R, and Pearlman MR (2005) The International Laser Ranging Service: Current status and future developments. *Advances in Space Research* 36(3): 327–332. http://dx.doi.org/10.1016/j.asr.2004.12.012.
Haas R and Wünsch J (2006) Sub-diurnal earth rotation variations from the VLBI CONT02 campaign. *Journal of Geodynamics* 41: 94–99.
Halley E (1695) Some account of the ancient state of the city of Palmyra; with short remarks on the inscriptions found there. *Philosophical Transactions of the Royal Society of London* 19: 160–175.
Hancock CM and Moore P (2007) On the usefulness of atmospheric and oceanic angular momentum in recovering polar motion and gravity field variations in a unified process. *Advances in Space Research* 39: 1648–1655. http://dx.doi.org/10.1016/j.asr.2007.02.051.
Haurwitz B and Cowley AD (1973) The diurnal and semidiurnal barometric oscillations, global distribution and annual variation. *Pure and Applied Geophysics* 102: 193–222.
Hefty J and Capitaine N (1990) The fortnightly and monthly zonal tides in the Earth's rotation from 1962 to 1988. *Geophysical Journal International* 103: 219–231.
Hefty J, Rothacher M, Springer T, Weber R, and Beutler G (2000) Analysis of the first year of Earth rotation parameters with a sub-daily resolution gained at the CODE processing center of the IGS. *Journal of Geodesy* 74: 479–487.
Hendon HH (1995) Length of day changes associated with the Madden-Julian oscillation. *Journal of Atmospheric Science* 52(13): 2373–2383.
Herring TA (1993) Diurnal and semidiurnal variations in Earth rotation. *Advances in Space Research* 13(11): 281–290.
Herring TA and Dong D (1991) Current and future accuracy of Earth rotation measurements. In: Carter WE (ed.) *Proceedings of the AGU Chapman Conference on Geodetic VLBI: Monitoring Global Change*, NOAA Technical Report NOS 137 NGS 49, pp 306–324. Washington, DC: NOAA.
Herring TA and Dong D (1994) Measurement of diurnal and semidiurnal rotational variations and tidal parameters of Earth. *Journal of Geophysical Research* 99(B9): 18051–18071.
Hide R (1969) Interaction between the Earth's liquid core and solid mantle. *Nature* 222: 1055–1056.
Hide R (1977) Towards a theory of irregular variations in the length of the day and core–mantle coupling. *Philosophical Transactions of the Royal Society of London A* 284: 547–554.
Hide R (1989) Fluctuations in the Earth's rotation and the topography of the core–mantle interface. *Philosophical Transactions of the Royal Society of London A* 328: 351–363.
Hide R (1993) Angular momentum transfer between the Earth's core and mantle. In: Le Mouël J-L, Smylie DE, and Herring T (eds.) *Dynamics of the Earth's Deep Interior and Earth Rotation*. *American Geophysical Union Geophysical Monograph Series*, vol. 72, pp. 109–112. Washington, DC, American Geophysical Union.
Hide R (1995a) The topographic torque on a bounding surface of a rotating gravitating fluid and the excitation by core motions of decadal fluctuations in the Earth's rotation. *Geophysical Research Letters* 22(8): 961–964.
Hide R (1995b) Reply to the comment by J. Bloxham and W. Kuang on a paper entitled "The topographic torque on a bounding surface of a rotating gravitating fluid and the excitation by core motions of decadal fluctuations in the Earth's rotation". *Geophysical Research Letters* 22(24): 3563–3565.
Hide R (1998) A note on topographic core–mantle coupling. *Physics of the Earth and Planetary Interiors* 109: 91–92.
Hide R, Birch NT, Morrison LV, Shea DJ, and White AA (1980) Atmospheric angular momentum fluctuations and changes in the length of the day. *Nature* 286: 114–117.
Hide R, Boggs DH, and Dickey JO (2000) Angular momentum fluctuations within the Earth's liquid core and torsional oscillations of the core–mantle system. *Geophysical Journal International* 143: 777–786.
Hide R, Boggs DH, Dickey JO, Dong D, Gross RS, and Jackson A (1996) Topographic core–mantle coupling and polar motion on decadal time scales. *Geophysical Journal International* 125: 599–607.
Hide R, Clayton RW, Hager BH, Spieth MA, and Voorhies CV (1993) Topographic core–mantle coupling and fluctuations in the Earth's rotation. In: Aki K and Dmowska R (eds.) *Relating Geophysical Structures and Processes: The Jeffreys Volume*. *American Geophysical Union Geophysical Monograph Series*, vol. 76, pp. 107–120. Washington, DC: American Geophysical Union.
Hide R and Dickey JO (1991) Earth's variable rotation. *Science* 253: 629–637.
Hinderer J, Legros H, Gire C, and Le Mouël J-L (1987) Geomagnetic secular variation, core motions and implications for the Earth's wobbles. *Physics of the Earth and Planetary Interiors* 49: 121–132.
Hinderer J, Legros H, Jault D, and Le Mouël J-L (1990) Core–mantle topographic torque: A spherical harmonic approach and implications for the excitation of the Earth's rotation by core motions. *Physics of the Earth and Planetary Interiors* 59: 329–341.
Hinnov LA and Wilson CR (1987) An estimate of the water storage contribution to the excitation of polar motion. *Geophysical Journal of the Royal Astronomical Society* 88: 437–459.
Hofmann-Wellenhof B, Lichtenegger H, and Wasle E (2008) *GNSS: Global Navigation Satellite Systems*. New York: Springer.
Holme R (1998a) Electromagnetic core–mantle coupling—I. Explaining decadal changes in the length of day. *Geophysical Journal International* 132: 167–180.
Holme R (1998b) Electromagnetic core–mantle coupling II: Probing deep mantle conductance. In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core-Mantle Boundary Region*. *American Geophysical Union Geodynamics Series*, vol. 28, pp. 139–151. Washington, DC: American Geophysical Union.
Holme R (2000) Electromagnetic core–mantle coupling III. Laterally varying mantle conductance. *Physics of the Earth and Planetary Interiors* 117: 329–344.
Holme R (2007) Large-scale flow in the core. In: Olson P (ed.) *Core Dynamics*. *Treatise on Geophysics*, vols. 107–130. Amsterdam: Elsevier.
Holme R and de Viron O (2005) Geomagnetic jerks and a high-resolution length-of-day profile for core studies. *Geophysical Journal International* 160: 435–439. http://dx.doi.org/10.1111/j.1365-246X.2004.02510.x.
Holme R and Olsen N (2006) Core surface flow modelling from high-resolution secular variation. *Geophysical Journal International* 166: 518–528. http://dx.doi.org/10.1111/j.1365-246X.2006.03033.x.

Holme R and Whaler KA (2001) Steady core flow in an azimuthally drifting reference frame. *Geophysical Journal International* 145: 560–569.
Höpfner J (1998) Seasonal variations in length of day and atmospheric angular momentum. *Geophysical Journal International* 135: 407–437.
Höpfner J (2000a) The International Latitude Service—A historical review, from the beginning to its foundation in 1899 and the period until 1922. *Surveys in Geophysics* 21: 521–566.
Höpfner J (2000b) Seasonal length-of-day changes and atmospheric angular momentum oscillations in their temporal variability. *Journal of Geodesy* 74: 335–358.
Höpfner J (2001) Atmospheric, oceanic, and hydrological contributions to seasonal variations in length of day. *Journal of Geodesy* 75: 137–150.
Höpfner J (2004) Low-frequency variations, Chandler and annual wobbles of polar motion as observed over one century. *Surveys in Geophysics* 25: 1–54.
Hough SS (1895) The oscillations of a rotating ellipsoidal shell containing fluid. *Philosophical Transactions of the Royal Society of London A* 186: 469–506.
Huang H-P, Weickmann KM, and Hsu CJ (2001) Trend in atmospheric angular momentum in a transient climate change simulation with greenhouse gas and aerosol forcing. *Journal of Climate* 14: 1525–1534.
Hulot G, Le Huy M, and Le Mouël J-L (1996) Influence of core flows on the decade variations of the polar motion. *Geophysical and Astrophysical Fluid Dynamics* 82: 35–67.
Jackson A (1997) Time-dependency of tangentially geostrophic core surface motions. *Physics of the Earth and Planetary Interiors* 103: 293–311.
Jackson A, Bloxham J, and Gubbins D (1993) Time-dependent flow at the core surface and conservation of angular momentum in the coupled core–mantle system. In: Le Mouël J-L, Smylie DE, and Herring T (eds.) *Dynamics of the Earth's Deep Interior and Earth Rotation*. *American Geophysical Union Geophysical Monograph Series*, vol. 72, pp. 97–107. Washington, DC: American Geophysical Union.
James TS and Ivins ER (1995) Present-day Antarctic ice mass changes and crustal motion. *Geophysical Research Letters* 22(8): 973–976.
James TS and Ivins ER (1997) Global geodetic signatures of the Antarctic ice sheet. *Journal of Geophysical Research* 102(B1): 605–633.
Jault D (2003) Electromagnetic and topographic coupling and LOD variations. In: Jones CA, Soward AM, and Zhang K (eds.) *Earth's core and lower mantle*, pp. 56–76. London: Taylor & Francis.
Jault D, Gire C, and Le Mouël J-L (1988) Westward drift, core motions and exchanges of angular momentum between core and mantle. *Nature* 333: 353–356.
Jault D, Hulot G, and Le Mouël J-L (1996) Mechanical core–mantle coupling and dynamo modelling. *Physics of the Earth and Planetary Interiors* 98: 187–191.
Jault D and Le Mouël J-L (1989) The topographic torque associated with tangentially geostrophic motion at the core surface and inferences on the flow inside the core. *Geophysical and Astrophysical Fluid Dynamics* 48: 273–296.
Jault D and Le Mouël J-L (1990) Core–mantle boundary shape: Constraints inferred from the pressure torque acting between the core and the mantle. *Geophysical Journal International* 101: 233–241.
Jault D and Le Mouël J-L (1991) Exchange of angular momentum between the core and the mantle. *Journal of Geomagnetism and Geoelectricity* 43: 111–129.
Jault D and Le Mouël J-L (1993) Circulation in the liquid core and coupling with the mantle. *Advances in Space Research* 13(11): 221–233.
Jault D and Le Mouël J-L (1999) Comment on "On the dynamics of topographical core-mantle coupling" by Weijia Kuang and Jeremy Bloxham. *Physics of the Earth and Planetary Interiors* 114: 211–215.
Jeffreys H (1928) Possible tidal effects on accurate time keeping. *Monthly Notices of the Royal Astronomical Society Geophysical Supplement* 2: 56–58.
Jeffreys H (1972) The variation of latitude. In: Melchior P and Yumi S (eds.) *Rotation of the Earth, Int. Astron. Union Symp. No. 48*, pp. 39–42. Dordrecht, Holland: D. Reidel.
Jin S, Chambers DP, and Tapley BD (2010) Hydrological and oceanic effects on polar motion from GRACE and models. *Journal of Geophysical Research* 115: B02403. http://dx.doi.org/10.1029/2009JB006635.
Jin S, Hassan AA, and Feng GP (2012) Assessment of terrestrial water contributions to polar motion from GRACE and hydrological models. *Journal of Geodynamics* 62: 40–48. http://dx.doi.org/10.1016/j.jog.2012.01.009.
Jin S, Zhang LJ, and Tapley BD (2011) The understanding of length-of-day variations from satellite gravity and laser ranging measurements. *Geophysical Journal International* 184: 651–660. http://dx.doi.org/10.1111/j.1365-246X.2010.04869.x.
Jochmann H (1989) Motion of the Earth's inner core and related variations of polar motion and the rotational velocity. *Astronomische Nachrichten* 310: 435–442.
Jochmann H (2009) Basic relations for studying the influence of geophysical processes on the Earth's rotation: The angular momentum approach. *Surveys in Geophysics* 30: 1–37. http://dx.doi.org/10.1007/s10712-009-9056-4.
Johnson TJ (2005) The interannual spectrum of the atmosphere and oceans. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 69–75. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.
Johnson TJ (2008) Recently discovered small polar motion loops and their atmospheric excitation mechanism. *Journal of Geophysical Research* 113: B04407. http://dx.doi.org/10.1029/2007JB005181.
Johnson TJ, Luzum BJ, and Ray JR (2005) Improved near-term Earth rotation predictions using atmospheric angular momentum analysis and forecasts. *Journal of Geodynamics* 39: 209–221.
Johnson TJ, Wilson CR, and Chao BF (1999) Oceanic angular momentum variability estimated from the Parallel Ocean Climate Model, 1988–1998. *Journal of Geophysical Research* 104(B11): 25183–25195.
Johnston P and Lambeck K (1999) Postglacial rebound and sea level contributions to changes in the geoid and the Earth's rotation axis. *Geophysical Journal International* 136: 537–558.
Jordi C, Morrison LV, Rosen RD, Salstein DA, and Rosselló G (1994) Fluctuations in the Earth's rotation since 1830 from high-resolution astronomical data. *Geophysical Journal International* 117: 811–818.
Kantha LH, Stewart JS, and Desai SD (1998) Long-period lunar fortnightly and monthly ocean tides. *Journal of Geophysical Research* 103(C6): 12639–12647.
Kaula WM (1966) *Theory of Satellite Geodesy*. Waltham, MA: Blaisdell.
King NE and Agnew DC (1991) How large is the retrograde annual wobble? *Geophysical Research Letters* 18(9): 1735–1738.
Kinoshita H, Nakajima K, Kubo Y, Nakagawa I, Sasao T, and Yokoyama K (1979) Note on nutation in ephemerides. *Publications of the International Latitude Observatory of Mizusawa* 12(2): 71–108.
Kolaczek B, Kosek W, and Schuh H (2000a) Short-period oscillations of Earth rotation. In: Dick S, McCarthy D, and Luzum B (eds.) *Polar Motion: Historical and Scientific Problems, IAU Colloq. 178, Astron. Soc. Pacific Conf. Ser.*, vol. 208, pp. 533–544. San Francisco: Astronomical Society of the Pacific.
Kolaczek B, Nastula J, and Salstein D (2003) El Niño-related variations in atmosphere-polar motion interactions. *Journal of Geodynamics* 36: 397–406.
Kolaczek B, Nuzhdina M, Nastula J, and Kosek W (2000b) El Niño impact on atmospheric polar motion excitation. *Journal of Geophysical Research* 105(B2): 3081–3087.
Kosek W, Nastula J, and Kolaczek B (1995) Variability of polar motion oscillations with periods from 20 to 150 days in 1979–1991. *Bulletin Géodesique* 69: 308–319.
Kouba J (2005) Comparison of polar motion with oceanic and atmospheric angular momentum time series for 2-day to Chandler periods. *Journal of Geodesy* 79: 33–42.
Kouba J and Vondrák J (2005) Comparison of length of day with oceanic and atmospheric angular momentum series. *Journal of Geodesy* 79: 256–268.
Kuang W and Bloxham J (1993) On the effect of boundary topography on flow in the Earth's core. *Geophysical and Astrophysical Fluid Dynamics* 72: 161–195.
Kuang W and Bloxham J (1997) On the dynamics of topographical core–mantle coupling. *Physics of the Earth and Planetary Interiors* 99: 289–294.
Kuang W and Chao B (2001) Topographic core–mantle coupling in geodynamo modeling. *Geophysical Research Letters* 28(9): 1871–1874.
Kuang W and Chao B (2003) Geodynamo modeling and core–mantle interactions. In: Dehant V, Creager KC, Karato S, and Zatman S (eds.) *Earth's Core: Dynamics, Structure, Rotation*. *American Geophysical Union Geodynamics Series*, vol. 31, pp. 193–212. Washington, DC: American Geophysical Union.
Kuehne J, Johnson S, and Wilson CR (1993) Atmospheric excitation of nonseasonal polar motion. *Journal of Geophysical Research* 98(B11): 19973–19978.
Kuehne J and Wilson CR (1991) Terrestrial water storage and polar motion. *Journal of Geophysical Research* 96(B3): 4337–4345.
Kuehne J, Wilson CR, and Johnson S (1996) Estimates of the Chandler wobble frequency and Q. *Journal of Geophysical Research* 101(B6): 13573–13579.
Lambeck K (1980) *The Earth's Variable Rotation: Geophysical Causes and Consequences*. New York: Cambridge University Press.
Lambeck K (1988) *Geophysical Geodesy: The Slow Deformations of the Earth*. New York: Oxford University Press.
Lambeck K and Hiopgood P (1981) The Earth's rotation and atmospheric circulation, from 1963 to 1973. *Geophysical Journal of the Royal Astronomical Society* 64: 67–89.
Lambert SB, Bizouard C, Dehant V, Lambert SB, Bizouard C, and Dehant V (2006) Rapid variations in polar motion during the 2005–2006 winter season. *Geophysical Research Letters* 33: L13303. http://dx.doi.org/10.1029/2006GL026422.
Landerer FW, Jungclaus JH, and Marotzke J (2007) Ocean bottom pressure changes lead to a decreasing length-of-day in a warming climate. *Geophysical Research Letters* 34: L06307. http://dx.doi.org/10.1029/2006GL029106.

Landerer FW, Jungclaus JH, and Marotzke J (2009) Long-term polar motion excited by ocean thermal expansion. *Geophysical Research Letters* 36: L17603. http://dx.doi.org/10.1029/2009GL039692.

Langley RB, King RW, and Shapiro II (1981a) Earth rotation from lunar laser ranging. *Journal of Geophysical Research* 86: 11913–11918.

Langley RB, King RW, Shapiro II, Rosen RD, and Salstein DA (1981b) Atmospheric angular momentum and the length of the day: A common fluctuation with a period near 50 days. *Nature* 294: 730–733.

Leick A (2003) *GPS Satellite Surveying*. New York: Wiley.

Lerch FJ, Nerem RS, Putney BH, et al. (1994) A geopotential model from satellite tracking, altimeter, and surface gravity data: GEM-T3. *Journal of Geophysical Research* 99(B2): 2815–2839.

Leuliette EW and Wahr JM (2002) Climate excitation of polar motion. In: Adám J and Schwarz K-P (eds.) *Vistas for Geodesy in the New Millennium, IAG Symposia*, vol. 125, pp. 428–433. Berlin: Springer.

Li Z (1985) Earth rotation from optical astrometry, 1962.0–1982.0. In: *Bureau International de l'Heure Annual Report for 1984*, pp D31–D63. Paris: Observatoire de Paris.

Li Z and Feissel M (1986) Determination of the Earth rotation parameters from optical astrometry observations, 1962.0–1982.0. *Bulletin Géodesique* 60: 15–28.

Liao D-C (2005) A brief review of atmospheric and oceanic excitation of the Chandler wobble. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 155–162. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.

Liao DC and Greiner-Mai H (1999) A new ΔLOD series in monthly intervals (1892.0–1997.0) and its comparison with other geophysical results. *Journal of Geodesy* 73: 466–477.

Liao D, Liao X, and Zhou Y (2003) Oceanic and atmospheric excitation of the Chandler wobble. *Geophysical Journal International* 152: 215–227.

Liao D, Liao X, Zhou Y, and Chen J (2004) Hydrological, atmospheric, and oceanic excitations of the Chandler wobble. *Chinese Astronomy and Astrophysics* 28: 212–221. http://dx.doi.org/10.1016/j.chinastron.2004.04.012.

Liao D, Zhou Y, and Liao X (2007) Comparison of wind contributions to Chandler wobble excitation. *Chinese Astronomy and Astrophysics* 31: 57–65. http://dx.doi.org/10.1016/j.chinastron.2007.01.005.

Lichten SM, Marcus SL, and Dickey JO (1992) Sub-daily resolution of Earth rotation variations with global positioning system measurements. *Geophysical Research Letters* 19(6): 537–540.

Liu L, Hsu H, and Grafarend EW (2007) Normal Morlet wavelet transform and its application to the Earth's polar motion. *Journal of Geophysical Research* 112: B08401. http://dx.doi.org/10.1029/2006JB004895.

Love JJ and Bloxham J (1994) Electromagnetic coupling and the toroidal magnetic field at the core–mantle boundary. *Geophysical Journal International* 117: 235–256.

Luzum BJ, Ray JR, Carter MS, and Josties FJ (2001) Recent improvements to IERS Bulletin A combination and prediction. *GPS Solutions* 4(3): 34–40.

Madden RA (1987) Relationships between changes in the length of day and the 40- to 50-day oscillations in the tropics. *Journal of Geophysical Research* 92: 8391–8399.

Madden RA and Julian PR (1971) Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *Journal of Atmospheric Science* 28: 702–708.

Madden RA and Julian PR (1972) Description of global-scale circulation cells in the tropics with a 40–50 day period. *Journal of Atmospheric Science* 29: 1109–1123.

Madden RA and Julian PR (1994) Observations of the 40-50-day tropical oscillation—A review. *Monthly Weather Review* 122: 814–837.

Madden RA and Julian PR (2005) Historical perspective. In: Lau WKM and Waliser DE (eds.) *Intraseasonal Variability in the Atmosphere—Ocean Climate System*, pp. 1–18. Berlin: Springer.

Malla RP, Wu SC, and Lichten SM (1993) Geocenter location and variations in Earth orientation using global positioning system measurements. *Journal of Geophysical Research* 98(B3): 4611–4617.

Mandea M, Holme R, Pais A, et al. (2010) Geomagnetic jerks: Rapid core field variations and core dynamics. *Space Science Reviews* 155: 147–175. http://dx.doi.org/10.1007/s11214-010-9663-x.

Marcus SL, Chao Y, Dickey JO, and Gegout P (1998) Detection and modeling of nontidal oceanic effects on Earth's rotation rate. *Science* 281: 1656–1659.

Marcus SL, de Viron O, and Dickey JO (2004) Atmospheric contributions to Earth nutation: Geodetic constraints and limitations of the torque approach. *Journal of the Atmospheric Sciences* 61: 352–356.

Marcus SL, de Viron O, and Dickey JO (2010) Interannual atmospheric torque and El Niño-Southern Oscillation: Where is the polar motion signal? *Journal of Geophysical Research* 115: B12409. http://dx.doi.org/10.1029/2010JB007524.

Marcus SL, Dickey JO, and de Viron O (2001) Links between intraseasonal (extended MJO) and ENSO timescales: Insights via geodetic and atmospheric analysis. *Geophysical Research Letters* 28(18): 3465–3468.

Marcus SL, Dickey JO, Fukumori I, and de Viron O (2012) Detection of the Earth rotation response to a rapid fluctuation of Southern Ocean circulation in November 2009. *Geophysical Research Letters* 39: L04605. http://dx.doi.org/10.1029/2011GL050671.

Markowitz W (1955) The annual variation in the rotation of the Earth, 1951–54. *The Astronomical Journal* 60: 171.

Markowitz W (1959) Variations in rotation of the Earth, results obtained with the dual-rate Moon camera and photographic zenith tubes. *The Astronomical Journal* 64: 106–113.

Markowitz W (1960) Latitude and longitude, and the secular motion of the pole. In: Runcorn SK (ed.) *Methods and Techniques in Geophysics*, pp. 325–361. New York: Interscience Publishers.

Markowitz W (1961) International determination of the total motion of the pole. *Bulletin Géodesique* 59: 29–41.

Marsh JG, Lerch FJ, Putney BH, et al. (1990) The GEM-T2 gravitational model. *Journal of Geophysical Research* 95(B13): 22043–22071.

Marsh JG, Lerch FJ, Putney BH, et al. (1991) Correction to "The GEM-T2 gravitational model". *Journal of Geophysical Research* 96(B10): 16651.

Martin CF (1969) *A Study of the Rate of Rotation of the Earth from Occultations of Stars by the Moon 1627–1860*. New Haven: Yale University, PhD Thesis, 144 pp.

Masaki Y (2008) Wind field differences between three meteorological reanalysis data sets detected by evaluating atmospheric excitation of Earth rotation. *Journal of Geophysical Research* 113: D07110. http://dx.doi.org/10.1029/2007JD008893.

Mathews PM and Bretagnon P (2003) Polar motions equivalent to high frequency nutations for a nonrigid Earth with anelastic mantle. *Astronomy and Astrophysics* 400: 1113–1128.

Mathews PM, Buffett BA, Herring TA, and Shapiro II (1991) Forced nutations of the Earth: Influence of inner core dynamics 2 Numerical results and comparisons. *Journal of Geophysical Research* 96: 8243–8257.

Mathews PM, Herring T, and Buffett BA (2002) Modeling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior. *Journal of Geophysical Research* 107(B4): 2068. http://dx.doi.org/10.1029/2001JB 000390.

Mathews PM and Lambert SB (2009) Effect of mantle and ocean tides on the Earth's rotation rate. *Astronomy and Astrophysics* 493: 325–330. http://dx.doi.org/10.1051/0004-6361:200810343.

Matsuyama I, Mitrovica JX, Daradich A, and Gomez N (2010) The rotational stability of a triaxial ice-age Earth. *Journal of Geophysical Research* 115: B05401. http://dx.doi.org/10.1029/2009JB006564.

McCarthy DD (1976) The determination of Universal Time at the U.S. Naval Observatory. *U.S. Naval Obs. Circ.*, No. 154.

McCarthy DD and Babcock AK (1986) The length of day since 1656. *Physics of the Earth and Planetary Interiors* 44: 281–292.

McCarthy D and Guinot B (2012) Time. In: Urban SE and Seidelmann PK (eds.) *Explanatory Supplement to the Astronomical Almanac*, 3rd edn., pp. 75–104. Mill Valley, CA: University Science Books.

McCarthy DD and Luzum BJ (1993) An analysis of tidal variations in the length of day. *Geophysical Journal International* 114: 341–346.

McCarthy DD and Luzum BJ (1996) Path of the mean rotational pole from 1899 to 1994. *Geophysical Journal International* 125: 623–629.

McCarthy D and Seidelmann PK (2009) *Time: From Earth Rotation to Atomic Physics*. Weinheim, Germany: Wiley-VCH.

Melchior P and Dejaiffe R (1969) Calcul des déclinaisons et mouvements propres des étoiles du Service International des Latitudes à partir des catalgues méridiens. *Annales Observatoire Royal de Belgique 3e serie* 10: 63–339.

Mendes Cerveira PJ, Boehm J, Schuh H, et al. (2009) Earth rotation observed by very long baseline interferometry and ring laser. *Pure and Applied Geophysics* 166: 1499–1517. http://dx.doi.org/10.1007/s00024-004-0487-z.

Merriam JB (1980) Zonal tides and changes in the length of day. *Geophysical Journal of the Royal Astronomical Society* 62: 551–561.

Merriam JB (1982) Meteorological excitation of the annual polar motion. *Geophysical Journal of the Royal Astronomical Society* 70: 41–56.

Merriam JB (1984) Tidal terms in Universal Time: Effects of zonal winds and mantle Q. *Journal of Geophysical Research* 89(B12): 10109–10114.

Merriam JB (1985) LAGEOS and UT measurements of long-period Earth tides and mantle Q. *Journal of Geophysical Research* 90(B11): 9423–9430.

Mitrovica JX and Forte AM (1997) Radial profile of mantle viscosity: Results from the joint inversion of convection and postglacial rebound observables. *Journal of Geophysical Research* 102(B2): 2751–2769.

Mitrovica JX, Latychev K, and Tamisiea ME (2007) Time variable gravity: Glacial isostatic adjustment. In: Herring T (ed.) *Geodesy. Treatise on Geophysics*, vol. 3, pp. 197–211. Amsterdam: Elsevier.

Mitrovica JX and Milne GA (1998) Glaciation-induced perturbations in the Earth's rotation: A new appraisal. *Journal of Geophysical Research* 103(B1): 985–1005.

Mitrovica JX and Peltier WR (1993) Present-day secular variations in the zonal harmonics of Earth's geopotential. *Journal of Geophysical Research* 98(B3): 4509–4526.

Mitrovica JX and Wahr J (2011) Ice age Earth rotation. *Annual Review of Earth and Planetary Sciences* 39: 577–616. http://dx.doi.org/10.1146/annurev-earth-040610-133404.

Mitrovica JX, Wahr J, Matsuyama I, and Paulson A (2005) The rotational stability of an ice-age Earth. *Geophysical Journal International* 161: 491–506.

Mitrovica JX, Wahr J, Matsuyama I, Paulson A, and Tamisiea ME (2006) Reanalysis of ancient eclipse, astronomic and geodetic data: A possible route to resolving the enigma of global sea-level rise. *Earth and Planetary Science Letters* 243: 390–399. http://dx.doi.org/10.1016/j.epsl.2005.12.029.

Moritz H and Mueller II (1988) *Earth Rotation: Theory and Observation*. New York: Ungar.

Morrison LV (1978) Catalogue of observations of occultations of stars by the Moon for the years 1943 to 1971. *Royal Greenwich Observatory Bulletin No. 183*.

Morrison LV (1979) Re-determination of the decade fluctuations in the rotation of the Earth in the period 1861–1978. *Geophysical Journal of the Royal Astronomical Society* 58: 349–360.

Morrison LV, Lukac MR, and Stephenson FR (1981) Catalogue of observations of occultations of stars by the Moon for the years 1623–1942 and solar eclipses for the years 1621–1806. *Royal Greenwich Observatory Bulletin No. 186*.

Morrison LV and Stephenson FR (2001) Historical eclipses and the variability of the Earth's rotation. *Journal of Geodynamics* 32: 247–265.

Morrison LV and Ward CG (1975) An analysis of the transits of Mercury: 1677–1973. *Monthly Notices of the Royal Astronomical Society* 173: 183–206.

Mound J (2005) Electromagnetic torques in the core and resonant excitation of decadal polar motion. *Geophysical Journal International* 160: 721–728.

Mound JE and Buffett BA (2003) Interannual oscillations in length of day: Implications for the structure of the mantle and core. *Journal of Geophysical Research* 108(B7): 2334. http://dx.doi.org/10.1029/2002JB002054.

Mound JE and Buffett BA (2005) Mechanisms of core–mantle angular momentum exchange and the observed spectral properties of torsional oscillations. *Journal of Geophysical Research* 110: B08103. http://dx.doi.org/10.1029/2004JB003555.

Mound JE and Buffett BA (2006) Detection of a gravitational oscillation in length-of-day. *Earth and Planetary Science Letters* 243: 383–389. http://dx.doi.org/10.1016/j.epsl.2006.01.043.

Mound J and Buffett B (2007) Viscosity of the Earth's fluid core and torsional oscillations. *Journal of Geophysical Research* 112: B05402. http://dx.doi.org/10.1029/2006JB004426.

Mulholland JD (1980) Scientific advances from ten years of lunar laser ranging. *Reviews of Geophysics and Space Physics* 18: 549–564.

Munk WH and MacDonald GJF (1960) *The Rotation of the Earth: A Geophysical Discussion*. New York: Cambridge University Press.

Naito I and Kikuchi N (1990) A seasonal budget of the Earth's axial angular momentum. *Geophysical Research Letters* 17: 631–634.

Naito I and Kikuchi N (1991) Reply to Rosen and Salstein's comment. *Geophysical Research Letters* 18: 1927–1928.

Nakada M (2000) Effect of the viscoelastic lithosphere on polar wander speed caused by the Late Pleistocene glacial cycles. *Geophysical Journal International* 143: 230–238.

Nakada M (2002) Polar wander caused by the Quaternary glacial cycles and fluid Love number. *Earth and Planetary Science Letters* 200: 159–166.

Nakada M (2006) Axial and equatorial rotations of the Earth's cores associated with the Quaternary ice age. *Physics of the Earth and Planetary Interiors* 154: 113–147. http://dx.doi.org/10.1016/j.pepi.2005.08.004.

Nakada M (2007) True polar wander associated with continental drift on a hypothetical Earth. *Earth, Planets and Space* 59: 513–522.

Nakada M (2008) Long-term true polar wander of the Earth including the effects of convective processes in the mantle and continental drift. *Geophysical Journal International* 175: 1235–1244. http://dx.doi.org/10.1111/j.1365-246X.2008.03935.x.

Nakada M (2009a) Earth's rotational variations by electromagnetic coupling due to core surface flow on a timescale of ~1 yr for geomagnetic jerk. *Geophysical Journal International* 179: 521–535. http://dx.doi.org/10.1111/j.1365-246X.2009.04256.x.

Nakada M (2009b) Polar wander of the Earth associated with the Quaternary glacial cycle on a convecting mantle. *Geophysical Journal International* 179: 569–578. http://dx.doi.org/10.1111/j.1365-246X.2009.04289.x.

Nakada M (2011) Earth's rotational variations due to rapid surface flows at both boundaries of the outer core. *Geophysical Journal International* 184: 235–246. http://dx.doi.org/10.1111/j.1365-246X.2010.04862.x.

Nakada M and Okuno J (2003) Perturbations of the Earth's rotation and their implications for the present-day mass balance of both polar ice caps. *Geophysical Journal International* 152: 124–138.

Nam YS and Dickman SR (1990) Effects of dynamic long-period ocean tides on changes in Earth's rotation rate. *Journal of Geophysical Research* 95(B5): 6751–6757.

Nastula J (1992) Short periodic variations in the Earth's rotation in the period 1984–1990. *Annales Geophysicae* 10: 441–448.

Nastula J (1995) Short periodic variations of polar motion and hemispheric atmospheric angular momentum excitation functions in the period 1984–1992. *Annales Geophysicae* 13: 217–225.

Nastula J (1997) The regional atmospheric contributions to the polar motion and EAAM excitation functions. In: Segawa J, Fujimoto H, and Okubo S (eds.) *Gravity, Geoid, and Marine Geodesy, IAG Symposia*, vol. 117, pp. 281–288. New York: Springer.

Nastula J, Gambis D, and Feissel M (1990) Correlated high-frequency variations in polar motion and length of the day in early 1988. *Annales Geophysicae* 8: 565–570.

Nastula J, Gross R, and Salstein DA (2012) Oceanic excitation of polar motion: Identification of specific oceanic areas important for polar motion excitation. *Journal of Geodynamics* 62: 16–23. http://dx.doi.org/10.1016/j.jog.2012.01.002.

Nastula J and Kolaczek B (2002) Seasonal oscillations in regional and global atmospheric excitation of polar motion. *Advances in Space Research* 30(2): 381–386.

Nastula J and Kolaczek B (2005) Analysis of hydrological excitation of polar motion. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 149–154. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.

Nastula J and Kolaczek B (2007) Spectral characteristic of polar motion in the 2005–2006 and 1999–2000 winter seasons. *Artificial Satellites* 42(1): 1–7. http://dx.doi.org/10.2478/v10018-007-0013-5.

Nastula J, Kolaczek B, Weber R, Schuh H, and Boehm J (2007a) Spectra of rapid oscillations of Earth rotation parameters determined during the CONT02 campaign. In: Tregoning P and Rizos C (eds.) *Dynamic Planet: Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools, IAG Symposia*, vol. 130, pp. 209–214. Berlin: Springer.

Nastula J, Kosek W, and Kolaczek B (1997) Analyses of zonal atmospheric excitation functions and their correlation with polar motion excitation functions. *Annales Geophysicae* 15: 1439–1446.

Nastula J and Ponte RM (1999) Further evidence for oceanic excitation of polar motion. *Geophysical Journal International* 139: 123–130.

Nastula J, Ponte RM, and Salstein DA (2000) Regional signals in atmospheric and oceanic excitation of polar motion. In: Dick S, McCarthy D, and Luzum B (eds.) *Polar Motion: Historical and Scientific Problems, IAU Colloq. 178, Astron. Soc. Pacific Conf. Ser.*, vol. 208, pp. 463–472. San Francisco: Astronomical Society of the Pacific.

Nastula J, Ponte RM, and Salstein DA (2002) Regional high-frequency signals in atmospheric and oceanic excitation of polar motion. *Advances in Space Research* 30(2): 369–374.

Nastula J, Ponte RM, and Salstein DA (2007b) Comparison of polar motion excitation series derived from GRACE and from analyses of geophysical fluids. *Geophysical Research Letters* 34: L11306. http://dx.doi.org/10.1029/2006GL028983.

Nastula J and Salstein D (1999) Regional atmospheric angular momentum contributions to polar motion excitation. *Journal of Geophysical Research* 104: 7347–7358.

Nastula J and Salstein DA (2012) Regional geophysical excitation functions of polar motion over land areas. In: Kenyon SC, Pacino MC, and Marti UJ (eds.) *Geodesy for Planet Earth, IAG Symposia*, vol. 136, pp. 491–497. Berlin: Springer.

Nastula J, Salstein D, and Kolaczek B (2005) Excitations of polar motion from an ensemble of global atmospheric models. In: Sansò F (ed.) *A Window on the Future of Geodesy, IAG Symposia*, vol. 128, pp. 597–600. Berlin: Springer.

Nastula J, Salstein D, and Kolaczek B (2009) Patterns of atmospheric excitation functions of polar motion from high-resolution regional sectors. *Journal of Geophysical Research* 114: B04407. http://dx.doi.org/10.1029/2008JB005605.

Nastula J, Salstein DA, and Ponte RM (2003) Empirical patterns of variability in atmospheric and oceanic excitation of polar motion. *Journal of Geodynamics* 36: 383–396.

Neef LJ and Matthes K (2012) Comparison of Earth rotation excitation in data-constrained and unconstrained atmosphere models. *Journal of Geophysical Research* 117: D02107. http://dx.doi.org/10.1029/2011JD016555.

Newhall XX, Williams JG, and Dickey JO (1988) Earth rotation from lunar laser ranging. In: Babcock AK and Wilkins GA (eds.) *Earth's Rotation and Reference Frame for Geodesy and Geodynamics*, pp. 159–164. Dordecht: Kluwer.

Nilsson T, Böhm J, Schuh H, Schreiber U, Gebauer A, and Klügel T (2012) Combining VLBI and ring laser observations for determination of high frequency Earth rotation variation. *Journal of Geodynamics* 62: 69–73. http://dx.doi.org/10.1016/j.jog.2012.02.002.

Okamoto I and Kikuchi N (1983) Low frequency variations of homogeneous ILS polar motion data. *Publications of the International Latitude Observatory of Mizusawa* 16: 35–40.

Olsen N and Mandea M (2007) Investigation of a secular variation impulse using satellite data: The 2003 geomagnetic jerk. *Earth and Planetary Science Letters* 255: 94–105. http://dx.doi.org/10.1016/j.epsl.2006.12.008.

Olsen N and Mandea M (2008) Rapidly changing flows in the Earth's core. *Nature Geoscience* 1: 390–394. http://dx.doi.org/10.1038/ngeo203.

Ooe M (1978) An optimal complex ARMA model of the Chandler wobble. *Geophysical Journal of the Royal Astronomical Society* 53: 445–457.

Pais A and Hulot G (2000) Length of day decade variations, torsional oscillations, and inner core superrotation: Evidence from recovered core surface zonal flows. *Physics of the Earth and Planetary Interiors* 118: 291–316.

Pais MA and Jault D (2008) Quasi-geostrophic flows responsible for the secular variation of the Earth's magnetic field. *Geophysical Journal International* 173: 421–443. http://dx.doi.org/10.1111/j.1365-246X.2008.03741.x.

Pais MA, Oliveira O, and Nogueira F (2004) Nonuniqueness of inverted core–mantle boundary flows and deviations from tangential geostrophy. *Journal of Geophysical Research* 109: B8105. http://dx.doi.org/10.1029/2004JB003012.

Pearlman MR, Degnan JJ, and Bosworth JM (2002) The International Laser Ranging Service. *Advances in Space Research* 30(2): 135–143.

Pearlman M, Noll C, Dunn P, et al. (2005) The International Laser Ranging Service and its support for IGGOS. *Journal of Geodynamics* 40: 470–478.

Peltier WR (1988) Global sea level and Earth rotation. *Science* 240: 895–901.

Peltier WR (2007) History of Earth rotation. In: Stevenson D (ed.) *Evolution of the Earth*. *Treatise on Geophysics*, vol. 9, pp. 243–293. Amsterdam: Elsevier.

Peltier WR and Drummond R (2010) Deepest mantle viscosity: Constraints from Earth rotation anomalies. *Geophysical Research Letters* 37: L12304. http://dx.doi.org/10.1029/2010GL043219.

Peltier WR and Jiang X (1996) Glacial isostatic adjustment and Earth rotation: Refined constraints on the viscosity of the deepest mantle. *Journal of Geophysical Research* 101(B2): 3269–3290.

Peltier WR and Luthcke SB (2009) On the origin of Earth rotation anomalies: New insights on the basis of both "paleogeodetic" data and Gravity Recovery and Climate Experiment (GRACE) data. *Journal of Geophysical Research* 114: B11405. http://dx.doi.org/10.1029/2009JB006352.

Peltier WR and Wu P (1983) Continental lithospheric thickness and deglaciation induced true polar wander. *Geophysical Research Letters* 10(3): 181–184.

Petit G and Luzum B (eds.) (2010) *IERS Conventions* (*2010*). IERS Tech. Note no. 36, 179 pp. Frankfurt, Germany: Bundesamts für Kartographie und Geodäsie.

Philander SG (1990) *El Niño, La Niña, and the Southern Oscillation*. San Diego, CA: Academic Press.

Ponsar S, Dehant V, Holme R, Jault D, Pais A, and Van Hoolst T (2003) The core and fluctuations in the Earth's rotation. In: Dehant V, Creager KC, Karato S, and Zatman S (eds.) *Earth's Core: Dynamics, Structure, Rotation*. *American Geophysical Union Geodynamics Series*, vol. 31, pp. 251–261. Washington, DC: American Geophysical Union.

Ponte RM (1997) Oceanic excitation of daily to seasonal signals in Earth rotation: Results from a constant-density numerical model. *Geophysical Journal International* 130: 469–474.

Ponte RM (2005) What do we know about low frequency signals in ocean angular momentum? In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 77–82. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.

Ponte RM and Ali AH (2002) Rapid ocean signals in polar motion and length of day. *Geophysical Research Letters* 29(15): 1711. http://dx.doi.org/10.1029/2002GL015312.

Ponte RM, Rajamony J, and Gregory JM (2002) Ocean angular momentum signals in a climate model and implications for Earth rotation. *Climate Dynamics* 19: 181–190. http://dx.doi.org/10.1007/s00382-001-0216-6.

Ponte RM and Stammer D (1999) Role of ocean currents and bottom pressure variability on seasonal polar motion. *Journal of Geophysical Research* 104: 23393–23409.

Ponte RM and Stammer D (2000) Global and regional axial ocean angular momentum signals and length-of-day variations (1985–1996). *Journal of Geophysical Research* 105: 17161–17171.

Ponte RM, Stammer D, and Marshall J (1998) Oceanic signals in observed motions of the Earth's pole of rotation. *Nature* 391: 476–479.

Ponte RM, Stammer D, and Wunsch C (2001) Improving ocean angular momentum estimates using a model constrained by data. *Geophysical Research Letters* 28: 1775–1778.

Räisänen J (2003) CO_2-induced changes in atmospheric angular momentum in CMIP2 experiments. *Journal of Climate* 16: 132–143.

Ray RD (1994) Tidal energy dissipation: Observations from astronomy, geodesy, and oceanography. In: Majumdar SK, Miller EW, Forbes GS, Schmalz RF, and Panah AA (eds.) *The Oceans: Physical-Chemical Dynamics and Human Impact*, pp. 171–185. Pittsburgh: Pennsylvania Academy of Science.

Ray RD, Bills BG, and Chao BF (1999) Lunar and solar torques on the oceanic tides. *Journal of Geophysical Research* 104(B8): 17653–17659.

Ray RD and Egbert GD (2012) Fortnightly Earth rotation, ocean tides and mantle anelasticity. *Geophysical Journal International* 189: 400–413. http://dx.doi.org/10.1111/j.1365-246X. 2012.05351.x.

Ray RD, Steinberg DJ, Chao BF, and Cartwright DE (1994) Diurnal and semidiurnal variations in the Earth's rotation rate induced by oceanic tides. *Science* 264: 830–832.

Ricard Y and Sabadini R (1990) Rotational instabilities of the Earth induced by mantle density anomalies. *Geophysical Research Letters* 17(5): 627–630.

Ricard Y, Sabadini R, and Spada G (1992) Isostatic deformations and polar wander induced by redistribution of mass within the Earth. *Journal of Geophysical Research* 97(B10): 14223–14236.

Ricard Y, Spada G, and Sabadini R (1993) Polar wandering of a dynamic Earth. *Geophysical Journal International* 113: 284–298.

Richards MA, Bunge H-P, Ricard Y, and Baumgardner JR (1999) Polar wandering in mantle convection models. *Geophysical Research Letters* 26(12): 1777–1780.

Richards MA, Ricard Y, Lithgow-Bertelloni C, Spada G, and Sabadini R (1997) An explanation for Earth's long-term rotational stability. *Science* 275: 372–375.

Roberts PH (1972) Electromagnetic core–mantle coupling. *Journal of Geomagnetism and Geoelectricity* 24: 231–259.

Roberts PH and Aurnou JM (2012) On the theory of core–mantle coupling. *Geophysical and Astrophysical Fluid Dynamics* 106(2): 157–230. http://dx.doi.org/10.1080/03091929.2011.589028.

Robertson DS (1991) Geophysical applications of very-long-baseline interferometry. *Reviews of Modern Physics* 63(4): 899–918.

Robertson DS, Ray JR, and Carter WE (1994) Tidal variations in UT1 observed with very long baseline interferometry. *Journal of Geophysical Research* 99(B1): 621–636.

Rochester MG (1960) Geomagnetic westward drift and irregularities in the Earth's rotation. *Philosophical Transactions of the Royal Society of London A* 252: 531–555.

Rochester MG (1962) Geomagnetic core–mantle coupling. *Journal of Geophysical Research* 67: 4833–4836.

Rochester MG (1984) Causes of fluctuations in the rotation of the Earth. *Philosophical Transactions of the Royal Society of London A* 313: 95–105.

Rochester MG and Smylie DE (1974) On changes in the trace of the Earth's inertia tensor. *Journal of Geophysical Research* 79(32): 4948–4951.

Roden RB (1963) Electromagnetic core–mantle coupling. *Geophysical Journal of the Royal Astronomical Society* 7: 361–374.

Rosen RD (1993) The axial momentum balance of Earth and its fluid envelope. *Surveys in Geophysics* 14: 1–29.

Rosen RD and Gutowski WJ Jr. (1992) Response of zonal winds and atmospheric angular momentum to a doubling of CO_2. *Journal of Climate* 5: 1391–1404.

Rosen RD and Salstein DA (1985) Contribution of stratospheric winds to annual and semi-annual fluctuations in atmospheric angular momentum and the length of day. *Journal of Geophysical Research* 90: 8033–8041.

Rosen RD and Salstein DA (1991) Comment on "A seasonal budget of the Earth's axial angular momentum" by Naito and Kikuchi. *Geophysical Research Letters* 18: 1925–1926.

Rosen RD and Salstein DA (2000) Multidecadal signals in the interannual variability of atmospheric angular momentum. *Climate Dynamics* 16: 693–700.

Rosen RD, Salstein DA, Eubanks TM, Dickey JO, and Steppe JA (1984) An El Niño signal in atmospheric angular momentum and Earth rotation. *Science* 225: 411–414.

Rosen RD, Salstein DA, and Wood TM (1990) Discrepancies in the Earth-atmosphere angular momentum budget. *Journal of Geophysical Research* 95(B1): 265–279.

Rothacher M, Beutler G, Weber R, and Hefty J (2001) High-frequency variations in Earth rotation from global positioning system data. *Journal of Geophysical Research* 106(B7): 13711–13738.

Rouby H, Greff-Lefftz M, and Besse J (2010) Mantle dynamics, geoid, inertia and TPW since 120 Myr. *Earth and Planetary Science Letters* 292: 301–311. http://dx.doi.org/10.1016/j.epsl.2010.01.033.

Roy K and Peltier WR (2011) GRACE era secular trends in Earth rotation parameters: A global scale impact of the global warming process? *Geophysical Research Letters* 38: L10306. http://dx.doi.org/10.1029/2011GL047282.

Rubincam DP (2003) Gravitational core–mantle coupling and the acceleration of the Earth. *Journal of Geophysical Research* 108(B7): 2338. http://dx.doi.org/10.1029/2002JB002132.

Sabadini R, Di Donato G, Vermeersen LLA, Devoti R, Luceri V, and Bianco G (2002) Ice mass loss in Antarctica and stiff lower mantle viscosity inferred from the long wavelength time dependent gravity field. *Geophysical Research Letters* 29(10): 1373. http://dx.doi.org/10.1029/2001GL014016.

Sabadini R and Peltier WR (1981) Pleistocene deglaciation and the Earth's rotation: Implications for mantle viscosity. *Geophysical Journal of the Royal Astronomical Society* 66: 553–578.

Sabadini R and Vermeersen BLA (2002) Long-term rotation instabilities of the Earth: A reanalysis. In: Mirtrovica JX and Vermeersen BLA (eds.) *Ice Sheets, Sea Level and the Dynamic Earth*. *American Geophysical Union Geodynamics Series*, vol. 29, pp. 51–67. Washington, DC: American Geophysical Union.

Sabadini R and Vermeersen B (2004) *Global Dynamics of the Earth: Applications of Normal Mode Relaxation Theory to Solid-Earth Geophysics*. Dordrecht, The Netherlands: Kluwer.

Sabadini R, Yuen DA, and Boschi E (1982) Polar wandering and the forced responses of a rotating, multilayered, viscoelastic planet. *Journal of Geophysical Research* 87(B4): 2885–2903.

Salstein DA (1993) Monitoring atmospheric winds and pressures for Earth orientation studies. *Advances in Space Research* 13(11): 175–184.

Salstein DA and Rosen RD (1986) Earth rotation as a proxy for interannual variability in atmospheric circulation, 1860-present. *Journal of Climate and Applied Meteorology* 25: 1870–1877.

Salstein DA and Rosen RD (1989) Regional contributions to the atmospheric excitation of rapid polar motions. *Journal of Geophysical Research* 94: 9971–9978.

Salstein DA and Rosen RD (1994) Topographic forcing of the atmosphere and a rapid change in the length of day. *Science* 264: 407–409.

Schaber K, Bunge H-P, Schuberth BSA, Malservisi R, and Horbach A (2009) Stability of the rotation axis in high-resolution mantle circulation models: Weak polar wander despite strong core heating. *Geochemistry, Geophysics, Geosystems* 10(11): Q11W04. http://dx.doi.org/10.1029/2009GC002541.

Schastok J, Soffel M, and Ruder H (1994) A contribution to the study of fortnightly and monthly zonal tides in UT1. *Astronomy and Astrophysics* 283: 650–654.

Schindelegger M, Böhm J, Salstein D, and Schuh H (2011) High-resolution atmospheric angular momentum functions related to Earth rotation parameters during CONT08. *Journal of Geodesy* 85: 425–433. http://dx.doi.org/10.0107/s00190-011-0458-y.

Schlüter W and Behrend D (2007) The International VLBI Service for Geodesy and Astrometry (IVS): Current capabilities and future prospects. *Journal of Geodesy* 81(6–8): 379–387. http://dx.doi.org/10.1007/s00190-006-0131-z.

Schlüter W, Himwich E, Nothnagel A, Vandenberg N, and Whitney A (2002) IVS and its important role in the maintenance of the global reference systems. *Advances in Space Research* 30(2): 145–150.

Schreiber KU, Klügel T, Velikoseltsev A, Schlüter W, Stedman GE, and Wells J-PR (2009) The large ring laser G for continuous Earth rotation monitoring. *Pure and Applied Geophysics* 166: 1485–1498. http://dx.doi.org/10.1007/s00024-004-0490-4.

Schreiber KU, Klügel T, Wells J-PR, Hurst RB, and Gebauer A (2011) How to detect the Chandler and the annual wobble of the Earth with a large ring laser gyroscope. *Physical Review Letters* 107: 173904. http://dx.doi.org/10.1103/Phys.RevLett.107.173904.

Schreiber KU, Velikoseltsev A, Rothacher M, Klügel T, Stedman GE, and Wiltshire DL (2004) Direct measurement of diurnal polar motion by ring laser gyroscopes. *Journal of Geophysical Research* 109: B06405. http://dx.doi.org/10.1029/2003JB002803.

Schuh H, Nagel S, and Seitz T (2001) Linear drift and periodic variations observed in long time series of polar motion. *Journal of Geodesy* 74: 701–710.

Schuh H, Richter B, and Nagel S (2000) Analysis of long time series of polar motion. In: Dick S, McCarthy D, and Luzum B (eds.) *Polar Motion: Historical and Scientific Problems, IAU Colloq. 178, Astron. Soc. Pacific Conf. Ser.*, vol. 208, pp. 321–331. San Francisco: Astronomical Society of the Pacific.

Schuh H and Schmitz-Hübsch H (2000) Short period variations in Earth rotation as seen by VLBI. *Surveys in Geophysics* 21: 499–520.

Segschneider J and Sündermann J (1997) Response of a global circulation model to real-time forcing and implications to Earth's rotation. *Journal of Physical Oceanography* 27: 2370–2380.

Seidelmann PK (1982) 1980 IAU theory of nutation: The final report of the IAU working group on nutations. *Celestial Mechanics* 27: 79–106.

Seidelmann PK, Guinot B, and Doggett LE (1992) Time. In: Seidelmann PK (ed.) *Explanatory Supplement to the Astronomical Almanac*. Mill Valley, CA: University Science Books, pp. 39–93.

Seiler U (1990) Variations of the angular momentum budget for tides of the present ocean. In: Brosche P and Sündermann J (eds.) *Earth's Rotation from Eons to Days*, pp. 81–94. New York: Springer.

Seiler U (1991) Periodic changes of the angular momentum budget due to the tides of the world ocean. *Journal of Geophysical Research* 96(B6): 10287–10300.

Seiler U and Wünsch J (1995) A refined model for the influence of ocean tides on UT1 and polar motion. *Astronomische Nachrichten* 316: 419–423.

Seitz F, Kirschner S, and Neubersch D (2012) Determination of the Earth's pole tide Love number *k*2 from observations of polar motion using an adaptive Kalman filter approach. *Journal of Geophysical Research* 117: B09403. http://dx.doi.org/10.1029/2012JB009296.

Seitz F and Kutterer H (2005) Sensitivity analysis of the non-linear Liouville equation. In: Sansò F (ed.) *A Window on the Future of Geodesy, IAG Symposia,* vol. 128, pp. 601–606. Berlin: Springer.

Seitz F and Schmidt M (2005) Atmospheric and oceanic contributions to Chandler wobble excitation determined by wavelet filtering. *Journal of Geophysical Research* 110: B11406. http://dx.doi.org/10.1029/2005JB003826.

Seitz F and Schuh H (2010) Earth rotation. In: Xu G (ed.) *Sciences of Geodesy I: Advances and Future Directions*, pp. 185–228. New York: Springer.

Seitz F, Stuck J, and Thomas M (2004) Consistent atmospheric and oceanic excitation of the Earth's free polar motion. *Geophysical Journal International* 157: 25–35.

Seitz F, Stuck J, and Thomas M (2005) White noise Chandler wobble excitation. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 15–21. Luxembourg, Cahiers du Centre Européen de Géodynamique et de Séismologie.

Seoane L, Biancale R, and Gambis D (2012) Agreement between Earth's rotation and mass displacement as detected by GRACE. *Journal of Geodynamics* 62: 49–55. http://dx.doi.org/10.1016/j.jog.2012.02.008.

Seoane L, Nastula J, Bizouard C, and Gambis D (2009) The use of gravimetric data from GRACE mission in the understanding of polar motion variations. *Geophysical Journal International* 178: 614–622. http://dx.doi.org/10.111/j.1365-246X.2009.04181.x.

Seoane L, Nastula J, Bizouard C, and Gambis D (2011) Hydrological excitation of polar motion derived from GRACE gravity field solutions. *International Journal of Geophysics* 2011: 174396. http://dx.doi.org/10.1155/2011/174396.

Shelus PJ (2001) Lunar laser ranging: Glorious past and a bright future. *Surveys in Geophysics* 22: 517–535.

Shirai T, Fukushima T, and Malkin Z (2005) Detection of phase disturbances of free core nutation of the Earth and their concurrence with geomagnetic jerks. *Earth, Planets and Space* 57: 151–155.

Sidorenkov N (2003) Influence of the atmospheric tides on the Earth rotation. *Celestial Mechanics and Dynamical Astronomy* 87: 27–38.

Simon JL, Bretagnon P, Chapront J, Chapront-Touzé M, Francou G, and Laskar J (1994) Numerical expressions for precession formulae and mean elements for the Moon and the planets. *Astronomy and Astrophysics* 282: 663–683.

Smith DE, Christodoulidis DC, Kolenkiewicz R, et al. (1985) A global geodetic reference frame from LAGEOS ranging (SL5.1AP). *Journal of Geophysical Research* 90: 9221–9233.

Smith ML and Dahlen FA (1981) The period and Q of the Chandler wobble. *Geophysical Journal of the Royal Astronomical Society* 64: 223–281.

Smith DE, Kolenkiewicz R, Dunn PJ, et al. (1990) Tectonic motion and deformation from satellite laser ranging to LAGEOS. *Journal of Geophysical Research* 95: 22013–22041.

Smith DE, Kolenkiewicz R, Nerem RS, et al. (1994) Contemporary global horizontal crustal motion. *Geophysical Journal International* 119: 511–520.

Soldati G and Spada G (1999) Large earthquakes and Earth rotation: The role of mantle relaxation. *Geophysical Research Letters* 26(7): 911–914.

Souriau A and Cazenave A (1985) Re-evaluation of the seismic excitation of the Chandler wobble from recent data. *Earth and Planetary Science Letters* 75: 410–416.

Sovers OJ, Fanselow JL, and Jacobs CS (1998) Astrometry and geodesy with radio interferometry: Experiments, models, results. *Reviews of Modern Physics* 70(4): 1393–1454.

Sovers OJ, Jacobs CS, and Gross RS (1993) Measuring rapid ocean tidal Earth orientation variations with very long baseline interferometry. *Journal of Geophysical Research* 98(B11): 19959–19971.

Spada G (1997) Why are earthquakes nudging the pole towards 140°E? *Geophysical Research Letters* 24(5): 539–542.

Spada G, Ricard Y, and Sabadini R (1992) Excitation of true polar wander by subduction. *Nature* 360: 452–454.
Spencer Jones H (1939) The rotation of the Earth and the secular acceleration of the Sun, Moon, and planets. *Monthly Notices of the Royal Astronomical Society Geophysical Supplement* 99: 541–558.
Stedman GE (1997) Ring-laser tests of fundamental physics and geophysics. *Reports on Progress in Physics* 60: 615–688.
Stefanick M (1982) Interannual atmospheric angular momentum variability 1963–1973 and the southern oscillation. *Journal of Geophysical Research* 87: 428–432.
Steigenberger P, Rothacher M, Dietrich R, Fritsche M, Rülke A, and Vey S (2006) Reprocessing of a global GPS network. *Journal of Geophysical Research* 111: B05402. http://dx.doi.org/10.1029/2005JB003747.
Steinberger BM and O'Connell RJ (1997) Changes of the Earth's rotation axis inferred from advection of mantle density heterogeneities. *Nature* 387: 169–173.
Steinberger B and O'Connell RJ (2002) The convective mantle flow signal in rates of true polar wander. In: Mirtrovica JX and Vermeersen BLA (eds.) *Ice Sheets, Sea Level and the Dynamic Earth*. *American Geophysical Union Geodynamics Series*, vol. 29, pp. 233–256. Washington, DC, American Geophysical Union.
Steinberger B and Torsvik TH (2010) Toward an explanation for the present and past locations of the poles. *Geochemistry, Geophysics, Geosystems* 11(6): Q06W06. http://dx.doi.org/10.1029/2009GC002889.
Stephenson FR (1997) *Historical Eclipses and Earth's Rotation*. New York: Cambridge University Press.
Stephenson FR (1998) Historical eclipses and the rotation of the Earth. *Physics and Chemistry of the Earth* 23(7–8): 715–724.
Stephenson FR and Morrison LV (1984) Long-term changes in the rotation of the Earth: 700 B.C. to A.D. 1980. *Philosophical Transactions of the Royal Society of London A* 313: 47–70.
Stephenson FR and Morrison LV (1985) Non-tidal changes in the length of the day: 700 BC to AD 1982. *Geophysical Surveys* 7: 201–210.
Stephenson FR and Morrison LV (1995) Long-term fluctuations in the Earth's rotation: 700 BC to AD 1990. *Philosophical Transactions of the Royal Society of London A* 351: 165–202.
Stephenson FR and Morrison LV (2005) Historical eclipses. In: Sterken C (ed.) *The Light-Time Effect in Astrophysics: Causes and Cures of the O—C Diagram*, *Astron. Soc. Pacific Conf. Ser.*, vol. 335, pp. 159–180. San Francisco: Astronomical Society of the Pacific.
Stewart DN, Busse FH, Whaler KA, and Gubbins D (1995) Geomagnetism, Earth rotation and the electrical conductivity of the lower mantle. *Physics of the Earth and Planetary Interiors* 92: 199–214.
Stieglitz TC and Dickman SR (1999) Refined correlations between atmospheric and rapid polar motion excitation. *Geophysical Journal International* 139: 115–122.
Stix M and Roberts PH (1984) Time-dependent electromagnetic core–mantle coupling. *Physics of the Earth and Planetary Interiors* 36: 49–60.
Stolz A, Bender PL, Faller JE, et al. (1976) Earth rotation measured by lunar laser ranging. *Science* 193: 997–999.
Stoyko N (1937) Sur la périodicité dans l'irrégularité de la rotation de la terre. *Comptes Rendus De L Academie Des Sciences* 205: 79–81.
Stuck J, Seitz F, and Thomas M (2005) Atmospheric forcing mechanisms of polar motion. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 127–133. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.
Su H, Neelin JD, and Meyerson JE (2005) Mechanisms for lagged atmospheric response to ENSO SST forcing. *Journal of Climate* 18: 4195–4215.
Tamisea ME, Mitrovica JX, Tromp J, and Milne GA (2002) Present-day secular variations in the low-degree harmonics of the geopotential: Sensitivity analysis on spherically symmetric Earth models. *Journal of Geophysical Research* 107(B12): 2378. http://dx.doi.org/10.1029/2001JB000696.
Tapley BD, Bettadpur S, Watkins M, and Reigber C (2004a) The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters* 31: L09607. http://dx.doi.org/10.1029/2004GL019920.
Tapley BD, Schutz BE, and Born GH (2004b) *Statistical Orbit Determination*. Burlington, MA: Elsevier.
Tapley BD, Schutz BE, and Eanes RJ (1985) Station coordinates, baselines, and Earth rotation from LAGEOS laser ranging: 1976–1984. *Journal of Geophysical Research* 90: 9235–9248.
Tapley BD, Schutz BE, Eanes RJ, Ries JC, and Watkins MM (1993) Lageos laser ranging contributions to geodynamics, geodesy, and orbital dynamics. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*. *American Geophysical Union Geodynamics Series*, vol. 24, pp. 147–173. Washington, DC: American Geophysical Union.
Tavernier G, Fagard H, Feissel-Vernier M, et al. (2005) The International DORIS Service (IDS). *Advances in Space Research* 36(3): 333–341.
Tavernier G, Fagard H, Feissel-Vernier M, et al. (2006) The International DORIS Service: Genesis and early achievements. *Journal of Geodesy* 80: 403–417. http://dx.doi.org/10.1007/s00190-006-0082-4.
Thaller D, Dill R, Krügel M, Steigenberger P, Rothacher M, and Tesmer V (2006) CONT02 analysis and combination of long EOP series. In: Flury J, Rummel R, Reigber C, Rothacher M, Boedecker G, and Schreiber U (eds.) *Observations of the Earth System from Space*, pp. 389–411. New York: Springer.
Thaller D, Krügel M, Rothacher M, Tesmer V, Schmid R, and Angermann D (2007) Combined Earth orientation parameters based on homogeneous and continuous VLBI and GPS data. *Journal of Geodesy* 81: 529–541. http://dx.doi.org/10.1007/s00190-006-0115-z.
Thaller D, Schmid R, Rothacher M, Tesmer V, and Angermann D (2005) Towards a rigorous combination of VLBI and GPS using the CONT02 campaign. In: Sansò F (ed.) *A Window on the Future of Geodesy, IAG Symposia,* vol. 128, pp. 576–581. Berlin: Springer.
Thomas M, Dobslaw H, Stuck J, and Seitz F (2005) The ocean's contribution to polar motion excitation – As many solutions as numerical models? In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 143–148. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.
Tisserand F (1891) *Traité de Mécanique Céleste*, vol. II. Paris: Gauthier-Villars.
Tosi N, Sabadini R, Marotta AM, and Vermeersen LLA (2005) Simultaneous inversion for the Earth's mantle viscosity and ice mass imbalance in Antarctica and Greenland. *Journal of Geophysical Research* 110: B07402. http://dx.doi.org/10.1029/2004JB003236.
Trenberth KE and Guillemot CJ (1994) The total mass of the atmosphere. *Journal of Geophysical Research* 99(D11): 23079–23088.
Trupin AS (1993) Effects of polar ice on the Earth's rotation and gravitational potential. *Geophysical Journal International* 113: 273–283.
Trupin AS, Meier MF, and Wahr JM (1992) Effect of melting glaciers on the Earth's rotation and gravitational field: 1965–1984. *Geophysical Journal International* 108: 1–15.
Tsai VC and Stevenson DJ (2007) Theoretical constraints on true polar wander. *Journal of Geophysical Research* 112: B05415. http://dx.doi.org/10.1029/2005JB003923.
Van Hoolst T and Dehant V (2002) Influence of triaxiality and second-order terms in flattenings on the rotation of terrestrial planets I Formalism and rotational normal modes. *Physics of the Earth and Planetary Interiors* 134: 17–33.
Vermeersen LLA, Fournier A, and Sabadini R (1997) Changes in rotation induced by Pleistocene ice masses with stratified analytical Earth models. *Journal of Geophysical Research* 102(B12): 27689–27702.
Vermeersen LLA and Sabadini R (1996) Significance of the fundamental mantle rotational relaxation mode in polar wander simulations. *Geophysical Journal International* 127: F5–F9.
Vermeersen LLA and Sabadini R (1999) Polar wander, sea-level variations and ice age cycles. *Surveys in Geophysics* 20: 415–440.
Vermeersen LLA, Sabadini R, and Spada G (1996) Compressible rotational deformation. *Geophysical Journal International* 126: 735–761.
Vermeersen LLA, Sabadini R, Spada G, and Vlaar NJ (1994) Mountain building and Earth rotation. *Geophysical Journal International* 117: 610–624.
Vermeersen LLA and Vlaar NJ (1993) Changes in the Earth's rotation by tectonic movements. *Geophysical Research Letters* 20(2): 81–84.
Vermeesen LLA, Sabadini R, Devoti R, et al. (1998) Mantle viscosity inferences from joint inversions of Pleistocene deglaciation-induced changes in geopotential with a new SLR analysis and polar wander. *Geophysical Research Letters* 25(23): 4261–4264.
Vicente RO and Wilson CR (1997) On the Variability of the chandler frequency. *Journal of Geophysical Research* 102(B9): 20439–20445.
Vicente RO and Wilson CR (2002) On long-period polar motion. *Journal of Geodesy* 76: 199–208.
Volland H (1988) *Atmospheric Tidal and Planetary Waves*. Dordrecht, The Netherlands: Kluwer.
Volland H (1997) Atmospheric tides. In: Wilhelm H, Zürn W, and Wenzel H-G (eds.) *Tidal Phenomena*. *Springer-Verlag Lecture Notes in Earth Sciences*, vol. 66, pp. 221–246. Berlin: Springer.
Vondrák J (1977) The rotation of the Earth between 1955.5 and 1976.5. *Studia Geophysica et Geoddaetica* 21: 107–117.
Vondrák J (1985) Long-period behaviour of polar motion between 1900.0 and 1984.0. *Annales Geophysicae* 3: 351–356.
Vondrák J (1991) Calculation of the new series of the Earth orientation parameters in the HIPPARCOS reference frame. *Bulletin of the Astronomical Institutes of Czechoslovakia* 42: 283–294.

Vondrák J (1994) Secular polar motion, crustal movements, and International Latitude Service observations. *Studia Geophysica et Geoddaetica* 38: 256–265.

Vondrák J (1999) Earth rotation parameters 1899.7–1992.0 after reanalysis within the Hipparcos frame. *Surveys in Geophysics* 20(2): 169–195.

Vondrák J, Feissel M, and Essaïfi N (1992) Expected accuracy of the 1900–1990 Earth orientation parameters in the Hipparcos reference frame. *Astronomy and Astrophysics* 262: 329–340.

Vondrák J, Pesek I, Ron C, and Cepek A (1998) *Earth Orientation Parameters 1899.7–1992.0 in the ICRS Based on the Hipparcos Reference Frame*. Ondrejov, Czech Republic: Astronomical Institute of the Academy of Sciences of the Czech Republic, Publication No. 87.

Vondrák J and Richter B (2004) International Earth Rotation and Reference Systems Service (IERS). *Journal of Geodesy* 77: 585–586.

Vondrák J, Ron C, and Pesek I (1997) Earth rotation in the Hipparcos reference frame. *Celestial Mechanics & Dynamical Astronomy* 66: 115–122.

Vondrák J, Ron C, Pesek I, and Cepek A (1995) New global solution of Earth orientation parameters from optical astrometry in 1900–1990. *Astronomy and Astrophysics* 297: 899–906.

Voorhies CV (1986) Steady flows at the top of Earth's core derived from geomagnetic field models. *Journal of Geophysical Research* 91(B12): 12444–12466.

Voorhies CV and Backus GE (1985) Steady flows at the top of the core from geomagnetic field models: The steady motions theorem. *Geophysical and Astrophysical Fluid Dynamics* 32: 163–173.

Wahr JM (1981) The forced nutations of an elliptical, rotating, elastic and oceanless earth. *Geophysical Journal of the Royal Astronomical Society* 64: 705–727.

Wahr JM (1982) The effects of the atmosphere and oceans on the Earth's wobble—I. Theory. *Geophysical Journal of the Royal Astronomical Society* 70: 349–372.

Wahr JM (1983) The effects of the atmosphere and oceans on the Earth's wobble and on the seasonal variations in the length of day—II. Results. *Geophysical Journal of the Royal Astronomical Society* 74: 451–487.

Wahr J (2005) Polar motion models: Angular momentum approach. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 89–102. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.

Wahr J and Bergen Z (1986) The effects of mantle anelasticity on nutations, Earth tides, and tidal variations in rotation rate. *Geophysical Journal of the Royal Astronomical Society* 87: 633–668.

Wahr JM, Sasao T, and Smith ML (1981) Effect of the fluid core on changes in the length of day due to long period tides. *Geophysical Journal of the Royal Astronomical Society* 64: 635–650.

Wardinski I, Holme R, Asari S, and Mandea M (2008) The 2003 geomagnetic jerk and its relation to the core surface flows. *Earth and Planetary Science Letters* 267: 468–481. http://dx.doi.org/10.1016/j.epsl.2007.12.008.

Wardinski I and Lesur V (2012) An extended version of the C3FM geomagnetic filed model: Application of a continuous frozen-flux constraint. *Geophysical Journal International* 189: 1409–1429. http://dx.doi.org/10.1111/j.1365-246X.2012.05384.x.

Watkins MM and Eanes RJ (1994) Diurnal and semidiurnal variations in Earth orientation determined from LAGEOS laser ranging. *Journal of Geophysical Research* 99(B9): 18073–18079.

Weber R, Nastula J, Kolaczek B, and Salstein DA (2002) Analysis of rapid variations of polar motion determined by GPS. In: Adám J and Schwarz K-P (eds.) *Vistas for Geodesy in the New Millennium, IAG Symposia*, vol. 125, pp. 445–450. Berlin: Springer.

Weis P (2006) Ocean Tides and the Earth's Rotation: Results of a High-resolving Ocean Model Forced by the Lunisolar Potential. PhD Thesis, 115 pp. Universität Hamburg.

Whaler KA (1980) Does the whole of the Earth's core convect? *Nature* 287: 528–530.

Whaler KA (1986) Geomagnetic evidence for fluid upwelling at the core–mantle boundary. *Geophysical Journal of the Royal Astronomical Society* 86: 563–588.

Whaler KA and Davis RG (1997) Probing the Earth's core with geomagnetism. In: Crossley DJ (ed.) *Earth's Deep Interior*, pp. 114–166. Amsterdam: Gordon and Breach.

Whaler KA and Holme R (2011) The axial dipole strength and flow in the outer core. *Physics of the Earth and Planetary Interiors* 188: 235–246. http://dx.doi.org/10.1016/j.pepi.2011.07.006.

Wicht J and Jault D (1999) Constraining electromagnetic core–mantle coupling. *Physics of the Earth and Planetary Interiors* 111: 161–177.

Wicht J and Jault D (2000) Electromagnetic core–mantle coupling for laterally varying mantle conductivity. *Journal of Geophysical Research* 105(B10): 23569–23578.

Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106(E11): 27933–27968.

Williams JG, Newhall XX, and Dickey JO (1993) Lunar laser ranging: Geophysical results and reference frames. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Earth Dynamics. American Geophysical Union Geodynamics Series*, vol. 24, pp. 83–88. Washington, DC: American Geophysical Union.

Williams JG, Turyshev SG, and Boggs DH (2009) Lunar laser ranging tests of the equivalence principle with the Earth and Moon. *International Journal of Modern Physics D* 18(7): 1129. http://dx.doi.org/10.1142/S021827180901500X.

Willis P, Fagard H, Ferrage P, et al. (2010) The International DORIS Service: Toward maturity. In: Willis P (ed.) DORIS: Scientific Applications in Geodesy and Geodynamics. *Advances in Space Research* 45(12): 1408–1420. http://dx.doi.org/10.1016/j.asr.2009.11.018.

Willis P, Jayles C, and Bar-Sever Y (2006) DORIS: From orbit determination for altimeter missions to geodesy. *Comptes Rendus Geoscience* 338: 968–979. http://dx.doi.org/10.1016/j.crte.2005.11.013.

Wilson CR (1993) Contributions of water mass redistribution to polar motion excitation. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Earth Dynamics. American Geophysical Union Geodynamics Series*, vol. 24, pp. 77–82. Washington, DC: American Geophysical Union.

Wilson CR and Chen J (2005) Estimating the period and Q of the Chandler wobble. In: Plag HP, Chao BF, Gross RS, and van Dam T (eds.) *Forcing of Polar Motion in the Chandler Frequency Band: A Contribution to Understanding Interannual Climate Change*, vol. 24, pp. 23–29. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie.

Wilson CR and Gabay S (1981) Excitation of the Earth's polar motion: A reassessment with new data. *Geophysical Research Letters* 8: 745–748.

Wilson CR and Haubrich RA (1976) Meteorological excitation of the Earth's wobble. *Geophysical Journal of the Royal Astronomical Society* 46: 707–743.

Wilson CR and Vicente RO (1980) An analysis of the homogeneous ILS polar motion series. *Geophysical Journal of the Royal Astronomical Society* 62: 605–616.

Wilson CR and Vicente RO (1990) Maximum likelihood estimates of polar motion parameters. In: McCarthy DD and Carter WE (eds.) *Variations in Earth Rotation. American Geophysical Union Geophysical Monograph Series*, vol. 59, pp. 151–155. Washington, DC: American Geophysical Union.

Winkelnkemper T, Seitz F, Min S-K, and Hense A (2009) Simulation of historic and future atmospheric angular momentum effects on length-of-day variations with GCMs. In: Sideris MG (ed.) *Observing Our Changing Earth, IAG Symposia*, vol. 133, pp. 447–454. Berlin: Springer.

Wu P and Peltier WR (1984) Pleistocene deglaciation and the Earth's rotation: A new analysis. *Geophysical Journal of the Royal Astronomical Society* 76: 753–791.

Wu B, Schuh H, and Bibo P (2003) New treatment of tidal braking of Earth rotation. *Journal of Geodynamics* 36: 515–521. http://dx.doi.org/10.1016/S0264-3707(03)00063-2.

Wünsch J (2000) Oceanic influence on the annual polar motion. *Journal of Geodynamics* 30: 389–399.

Wünsch J (2002) Oceanic and soil moisture contributions to seasonal polar motion. *Journal of Geodynamics* 33: 269–280.

Wünsch J and Busshoff J (1992) Improved observations of periodic UT1 variations caused by ocean tides. *Astronomy and Astrophysics* 266: 588–591.

Xu S, Crossley D, and Szeto AMK (2000) Variations in length of day and inner core differential rotation from gravitational coupling. *Physics of the Earth and Planetary Interiors* 117: 95–110.

Yan H and Chao BF (2012) Effect of global mass conservation among geophysical fluids on the seasonal length of day variation. *Journal of Geophysical Research* 117: B02401. http://dx.doi.org/10.1029/2011JB008788.

Yan H, Zhong M, Zhu Y, Liu L, and Cao X (2006) Nontidal oceanic contribution to length-of-day changes estimated from two ocean models during 1992–2001. *Journal of Geophysical Research* 111: B02410. http://dx.doi.org/10.1029/2004JB003538.

Yoder CF (1995) Astrometric and geodetic properties of Earth and the solar system. In: Ahrens TJ (ed.) *Global Earth Physics: A Handbook of Physical Constants. American Geophysical Union Reference Shelf 1*, pp. 1–31. Washington, DC: American Geophysical Union.

Yoder CF and Standish EM (1997) Martian precession and rotation from Viking lander range data. *Journal of Geophysical Research* 102(E2): 4065–4080.

Yoder CF, Williams JG, and Parke ME (1981) Tidal variations of Earth rotation. *Journal of Geophysical Research* 86: 881–891.

Yu N, Zheng D, and Wu H (1999) Contribution of new AAM data source to ΔLOD excitation. *Journal of Geodesy* 73: 385–390.

Yuen DA, Sabadini R, and Boschi EV (1982) Viscosity of the lower mantle as inferred from rotational data. *Journal of Geophysical Research* 87(B13): 10745–10762.

Yumi S and Yokoyama K (1980) *Results of the International Latitude Service in a Homogeneous System, 1899.9–1979.0*. Mizusawa, Japan: Publication of the Central Bureau of the International Polar Motion Service and the International Latitude Observatory of Mizusawa.

Zatman S (2003) Decadal oscillations of the Earth's core, angular momentum exchange, and inner core rotation. In: Dehant V, Creager KC, Karato S, and Zatman S (eds.) *Earth's Core: Dynamics, Structure, Rotation*. *American Geophysical Union Geodynamics Series*, vol. 31, pp. 233–240. Washington, DC: American Geophysical Union.

Zatman S and Bloxham J (1997) Torsional oscillations and the magnetic field within the Earth's core. *Nature* 388: 760–763.

Zatman S and Bloxham J (1998) A one-dimensional map of Bs from torsional oscillations of the Earth's core. In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region*. *American Geophysical Union Geodynamics Series*, vol. 28, pp. 183–196. Washington, DC: American Geophysical Union.

Zatman S and Bloxham J (1999) On the dynamical implications of models of *Bs* in the Earth's core. *Geophysical Journal International* 138: 679–686.

Zhao M and Dong D (1988) A new research for the secular polar motion in this century. In: Babcock AK and Wilkins GA (eds.) *The Earth's Rotation and Reference Frames for Geodesy and Geodynamics, Int. Astron. Union Symp. No. 128*, pp. 385–392. Dordrecht, Holland: D. Reidel.

Zharov VE and Gambis D (1996) Atmospheric tides and rotation of the Earth. *Journal of Geodesy* 70: 321–326.

Zheng D, Ding X, Zhou Y, and Chen Y (2003) Earth rotation and ENSO events: Combined excitation of interannual LOD variations by multiscale atmospheric oscillations. *Global and Planetary Change* 36: 89–97.

Zhong M, Naito I, and Kitoh A (2003) Atmospheric, hydrological, and ocean current contributions to Earth's annual wobble and length-of-day signals based on output from a climate model. *Journal of Geophysical Research* 108(B1): 2057. http://dx.doi.org/10.1029/2001JB000457.

Zhong M, Yan H, Wu X, Duan J, and Zhu Y (2006) Non-tidal oceanic contribution to polar wobble estimated from two oceanic assimilation data sets. *Journal of Geodynamics* 41: 147–154.

Zhou YH, Chen JL, Liao XH, and Wilson CR (2005) Oceanic excitations on polar motion: A cross comparison among models. *Geophysical Journal International* 162: 390–398.

Zhou YH, Chen JL, and Salstein DA (2008) Tropospheric and stratospheric wind contributions to Earth's variable rotation from NCEP/NCAR reanalyses (2000–2005). *Geophysical Journal International* 174: 453–463. http://dx.doi.org/10.1111/j.1365-246X.2008.03843.x.

Zhou YH, Salstein DA, and Chen JL (2006) Revised atmospheric excitation function series related to Earth's variable rotation under consideration of surface topography. *Journal of Geophysical Research* 111: D12108. http://dx.doi.org/10.1029/2005JD006608.

Zhou J, Sun W, Sun H, and Xu J (2013) Reformulation of co-seismic polar motion excitation and low degree gravity changes: Applied to the 2011 Tohoku-Oki earthquake (Mw9.0). *Journal of Geodynamics* 63: 20–26. http://dx.doi.org/10.1016/j.jog.2012.09.004.

Zhou YH, Zheng DW, and Liao XH (2001) Wavelet analysis of interannual LOD, AAM, and ENSO: 1997–98 El Niño and 1998–99 La Niña signals. *Journal of Geodesy* 75: 164–168.

Zhou Y, Zheng D, Zhao M, and Chao BF (1998) Interannual polar motion with relation to the North Atlantic Oscillation. *Global and Planetary Change* 18: 79–84.

Zotov L and Bizouard C (2012) On modulations of the Chandler wobble excitation. *Journal of Geodynamics* 62: 30–34. http://dx.doi.org/10.1016/j.jog.2012.03.010.

This page intentionally left blank

3.10 Earth Rotation Variations

V Dehant, Royal Observatory of Belgium, Brussels, Belgium
PM Mathews, University of Madras, Chennai, India

© 2015 Elsevier B.V. All rights reserved.

3.10.1 Introduction–Concepts–Overview 267
3.10.1.1 Manifestations of Variations in Earth Rotation and Orientation 267
3.10.1.2 Free Rotational Modes 269
3.10.1.3 Forced Motions 269
3.10.1.4 Earth's Response to Gravitational Torques 270
3.10.1.5 Low-Frequency Nutations 270
3.10.1.6 Gravitational Perturbations and Displacement Fields 271
3.10.1.7 Gravitational Torque on the Earth 271
3.10.1.8 Earth Response to the Torque 271
3.10.1.9 Role of Deformations in Earth Rotation 272
3.10.1.10 Interplay of Various Phenomena 272
3.10.1.11 Precision of Observations: Challenge to Theory 272
3.10.2 Earth Orientation/Rotation Variables: Reference Frames 273
3.10.2.1 Ecliptic Celestial Reference Frame 273
3.10.2.1.1 Precession 273
3.10.2.1.2 Nutation 273
3.10.2.2 Geocentric Celestial Reference Frame 273
3.10.2.3 International Terrestrial Reference Frame 274
3.10.2.4 Principal Axis Frame 274
3.10.2.5 The CIP and CIO 274
3.10.2.5.1 Nutation and polar motion in the conventions: CIP 275
3.10.2.5.2 LOD variations in the conventions: CIO 275
3.10.3 Equations of Rotational Motion 275
3.10.3.1 Equation of Motion in a Celestial Frame 275
3.10.3.2 Equation of Motion in a Terrestrial Frame 275
3.10.3.2.1 Euler equations for a rigid body 276
3.10.3.2.2 Axially symmetrical ellipsoidal case: Wobble motion 276
3.10.4 The Tidal Potential and Torque 276
3.10.4.1 Potential, Acceleration, and Torque 277
3.10.4.2 Torque Components in Terrestrial Frame 277
3.10.4.3 Expansion of the Potential: Tidal Spectrum 278
3.10.4.4 Degree 2 Potential and Torque: Spectra 279
3.10.4.4.1 Equatorial components of torque 279
3.10.4.4.2 Axial component of the torque 280
3.10.4.5 Terms of General Degree and Order in the Torque 280
3.10.5 Torque in Celestial Frame: Nutation–Precession in a Simple Model 281
3.10.5.1 Torque Components in Celestial Frame 281
3.10.5.2 Simple Model: Sun in a Circular Orbit 281
3.10.5.3 The Torque in the Simple Model 281
3.10.5.4 Nutation and Precession in the Model 281
3.10.6 Elliptical Motions: Prograde and Retrograde Circular Components 282
3.10.6.1 The Spectrum and the Fundamental Arguments of Nutation 282
3.10.6.2 Resolution of Elliptical Motions into Pairs of Circular Motions 283
3.10.6.3 Nutation Series 284
3.10.7 Kinematic Relations between the Nutation of the Figure Axis and the Wobble 285
3.10.7.1 Kinematic Relations in the Frequency Domain 285
3.10.7.2 Implications of the Kinematic Relations 286
3.10.7.3 Transfer Function 286
3.10.7.4 Relations Connecting the Nutations of Different Axes: Oppolzer Terms 287
3.10.8 Rigid Earth Nutation 288
3.10.8.1 Hamiltonian Approach 288
3.10.8.2 Torque Approach 290

3.10.8.2.1	Axially symmetrical Earth: simplified treatment	290
3.10.8.2.2	Rigorous treatment of the general case	290
3.10.9	**Axially Symmetrical Ellipsoidal Nonrigid Earth: Torque Equations and Solutions**	291
3.10.9.1	Deformable Wholly Solid Earth	291
3.10.9.2	Two-Layer Earth, with Mantle and Fluid Core	292
3.10.9.2.1	Equations of motion	292
3.10.9.2.2	Wobble normal modes	293
3.10.9.2.3	Solution of the wobble equations: resonances and nutation amplitudes	293
3.10.9.3	Coupling of the Core and the Mantle at the CMB	293
3.10.9.4	Anelasticity and Ocean Tide Effects	294
3.10.9.5	Inclusion of SIC	294
3.10.9.5.1	Solution of the equations for the three-layer Earth	294
3.10.9.6	Confronting Theory with Observations	295
3.10.10	**Nutation–Precession from the Displacement Field**	297
3.10.10.1	Displacement Field Approach	297
3.10.10.1.1	Reference system used	297
3.10.10.1.2	Earth model parameters and unknowns used	297
3.10.10.1.3	Basic equations	298
3.10.10.1.4	Generalized spherical harmonics (GSH) expansions: radial functions	298
3.10.10.1.5	Equations for radial functions	299
3.10.10.1.6	Solutions of the radial equations	300
3.10.11	**Atmospheric Tides and Nontidal Effects from Surficial Fluids**	301
3.10.12	**New Conventions for Earth Rotation Variations**	302
3.10.12.1	CIP and Relation of its Motions to the Nutation and Wobble	302
3.10.12.2	Nonrotating Origin	302
3.10.12.3	Definition of Universal Time	302
3.10.12.4	Transformation between ICRF and ITRF	302
References		304

Nomenclature

$\boldsymbol{\Omega_0}$	Mean angular velocity vector
Ω_0	Mean Earth angular velocity = 2π radian per sidereal day; $= 7.292115 \times 10^{-5}$ rad $s^{-1} = 1$ cpsd
a	Mean radius of the Earth = 6371 km
a_e	Equatorial radius of the Earth = 6378 km
θ	Colatitude
λ	Longitude
T_d	Period of the relative orbital motion of the Earth and the Sun in solar days: $T_d = 365.2422$ days
H_d	Dynamical ellipticity of the Earth (astronomy community): $H_d = \frac{C-A}{C} \approx \frac{C-A}{C}, \quad H_d = \frac{e}{(1+e)}$
e	Dynamical ellipticity of the Earth (geodesy community): $e = \frac{C-\bar{A}}{A} \approx \frac{C-A}{C}, \quad e = \frac{H_d}{1-H_d}$
e'	Triaxiality coefficient of the Earth: $e' = \frac{B-A}{A+B} = \frac{B-A}{2A}$
A, B	Equatorial moments of inertia: $A \approx B$
$\bar{A}$	Mean equatorial moment of inertia: $\bar{A} = \frac{A+B}{2}$
C	Polar moment of inertia: $C > B > A$
ϵ	Obliquity
ϵ_0	Mean obliquity at J2000
$\psi_A(t)$	Precession in longitude of the figure axis: more precisely of the CIP
$\dot{\psi}_A$	Precession rate in longitude of the figure axis: more precisely of the CIP
$\Delta\psi$	Nutation in longitude of the figure axis: more precisely of the CIP
$\epsilon_A(t)$	Mean obliquity of date
$\dot{\epsilon}_A$	Precession rate in obliquity of the figure axis: more precisely of the CIP
$\Delta\epsilon$	Nutation in obliquity of the figure axis: more precisely of the CIP
X, Y	Coordinates of the CIP in the GCRF: $X + iY = \Delta\psi \sin\epsilon + i\Delta\epsilon$
$\tilde{\eta}$	Nutation of the CIP: $\tilde{\eta} = \Delta\psi \sin\epsilon + i\Delta\epsilon = X + iY$
$\tilde{\eta}_n(t)$	Circular nutation component: $\tilde{\eta}_n(t) = -i\tilde{\eta}(\omega_n)e^{i\Xi_n(t)}$
ω_n	Frequency of nutation in the CRF: $\Xi_n(t) = \omega_n t + \mathcal{X}_n$
ν	Frequency of nutation in the CRF in cpsd: $\omega_n = \nu\Omega_0$
$\mathcal{X}_n$	Phase of component of frequency ω_n of $\tilde{\eta}$
$\tilde{\eta}(\omega_n)$	Amplitude of component of frequency ω_n of $\tilde{\eta}$
$\tilde{\eta}_{Rig}(\omega_n)$	Rigid Earth nutation amplitude of component of frequency ω_n
$\Xi(t)$ or $\Xi_n(t)$	Nutation argument: $\Xi(t) = n_1 l + n_2 l' + n_3 F + n_4 D + n_5 \Omega + \sum_{i=1}^{9} n_i \lambda_i$

ℓ	Mean anomaly of the Moon
ℓ'	Mean anomaly of the Sun
F	One of the fundamental nutation arguments: $F = L - \Omega$
L	Mean longitude of the Moon
D	Mean elongation of the Moon from the Sun
$\boldsymbol{\Omega}$	Mean longitude of the Moon's ascending node
$\boldsymbol{\lambda}_i$	Mean longitude of the planet i
A, A', A''	Amplitudes of the nutation in longitude: $\Delta\psi(t) = \sum_{\mathrm{n}}[(A_{\mathrm{n}} + A'_{\mathrm{n}}t)\sin \Xi_{\mathrm{n}}(t) + A''_{\mathrm{n}}\cos \Xi_{\mathrm{n}}(t)]$
B, B', B''	Amplitudes of the nutation in obliquity: $\Delta\varepsilon(t) = \sum_{\omega}[(B_{\mathrm{n}} + B'_{\mathrm{n}}t)\cos \Xi_{\mathrm{n}}(t) + B''_{\mathrm{n}}\sin \Xi_{\mathrm{n}}(t)]$
$X_{\mathrm{s}}, X_{\mathrm{c}}$	Sine and cosine components of X: $X = Y_{\mathrm{s}}\sin \Xi + X_{\mathrm{c}}\cos \Xi$
$Y_{\mathrm{s}}, Y_{\mathrm{c}}$	Sine and cosine components of Y: $Y = Y_{\mathrm{c}}\cos \Xi + Y_{\mathrm{s}}\sin \Xi$
$\boldsymbol{\eta}^{p,\mathrm{ip}}$	Amplitude of the prograde in-phase component: $\eta^{p,\mathrm{ip}} = \frac{1}{2}\left(\frac{\omega_{\mathrm{n}}}{\lvert\omega_{\mathrm{n}}\rvert}X_{\mathrm{s}} - Y_{\mathrm{c}}\right)$
$\boldsymbol{\eta}^{p,\mathrm{op}}$	Amplitude of the prograde out-of-phase component: $\eta^{p,\mathrm{op}} = \frac{1}{2}\left(X_{\mathrm{c}} + \frac{\omega_{\mathrm{n}}}{\lvert\omega_{\mathrm{n}}\rvert}Y_{\mathrm{s}}\right)$
$\boldsymbol{\eta}^{r,\mathrm{ip}}$	Amplitude of the retrograde in-phase component: $\eta^{r,\mathrm{ip}} = -\frac{1}{2}\left(\frac{\omega_{\mathrm{n}}}{\lvert\omega_{\mathrm{n}}\rvert}X_{\mathrm{s}} + Y_{\mathrm{c}}\right)$
$\boldsymbol{\eta}^{r,\mathrm{op}}$	Amplitude of the retrograde out-of-phase component: $\eta^{r,\mathrm{op}} = \frac{1}{2}\left(X_{\mathrm{c}} - \frac{\omega_{\mathrm{n}}}{\lvert\omega_{\mathrm{n}}\rvert}Y_{\mathrm{s}}\right)$
$\boldsymbol{\Omega}$	Earth's angular velocity vector: $\boldsymbol{\Omega} = (\Omega_x, \Omega_y, \Omega_z)$
$\mathbf{m}$	Wobble: $\boldsymbol{\Omega} = \boldsymbol{\Omega}_0 + \boldsymbol{\Omega}_0\mathbf{m}$
m_z	Wobble axial component: $m_z = \frac{\Omega_z}{\Omega_0} - 1 = \frac{-\Delta\mathrm{LOD}}{\mathrm{LOD}}$
m_x, m_y	Wobble equatorial components: $m_x = \frac{\Omega_x}{\Omega_0}$ and $m_y = \frac{\Omega_y}{\Omega_0}$
$\tilde{m}$	Complex sum of the two previously mentioned components: $\tilde{m} = m_x + \mathrm{i}m_y$ $\tilde{m} = \tilde{m}(\omega_{\mathrm{w}})\mathrm{e}^{\mathrm{i}(\omega_{\mathrm{w}}t + \mathcal{X}_{\mathrm{w}})}$ $\Omega_0\tilde{m} = i\dot{\tilde{\eta}}\mathrm{e}^{-\mathrm{i}\Omega_0(t-t_0)}$
$\boldsymbol{\omega}_{\mathrm{w}} = \boldsymbol{\omega}$	Frequency of the equatorial wobble: $\omega_{\mathrm{w}} = \omega_{\mathrm{n}} - \Omega_0$
$\boldsymbol{\sigma}$	Frequency of the equatorial wobble in cpsd: $\sigma = \frac{\omega_{\mathrm{w}}}{\Omega_0} = \nu - 1$
$\boldsymbol{\Theta}_{\mathrm{w}}(t) = \boldsymbol{\omega}t + \mathcal{X}_{\mathrm{w}}$	Argument of the equatorial wobble
$\mathcal{X}_{\mathrm{w}}$	Phase of the equatorial wobble
$\tilde{m}(\boldsymbol{\omega}_{\mathrm{w}})$	Amplitude of the component of frequency ω_{w} of $\tilde{m}$: $\tilde{m}(\omega_{\mathrm{w}}) = -\frac{\omega_{\mathrm{n}}}{\Omega_0}\tilde{\eta}(\omega_{\mathrm{n}}) = \frac{\omega_{\mathrm{n}}}{\Omega_0}\tilde{p}(\omega_{\mathrm{p}})$
$\mathbf{H}$	Angular momentum vector
$\boldsymbol{\Gamma}$	Torque vector
(H_x, H_y, H_z)	Components of $\mathbf{H}$ in space
$(\Gamma_x, \Gamma_y, \Gamma_z)$	Components of $\boldsymbol{\Gamma}$ in space
X_H, Y_H	Normalized components of $\mathbf{H}$ in space: $X_{\mathrm{H}} = \frac{H_X}{C\Omega_0}$, $Y_{\mathrm{H}} = \frac{H_Y}{C\Omega_0}$
$\boldsymbol{\Delta\psi}_{\mathbf{H}}$	Nutation in longitude of the angular momentum axis
$\boldsymbol{\Delta\varepsilon}_{\mathbf{H}}$	Nutation in obliquity of the angular momentum axis
$\boldsymbol{\psi}_{\mathrm{A,H}}(t)$	Precession in longitude of the angular momentum axis
$\dot{\boldsymbol{\psi}}_{\mathrm{A,\,H}}$	Precession rate in longitude of the angular momentum axis
(H_x, H_y, H_z)	Components of $\mathbf{H}$ in TRF
$(\Gamma_x, \Gamma_y, \Gamma_z)$	Components of $\boldsymbol{\Gamma}$ in TRF
$\tilde{\Gamma}$	Complex sum of the first two components previously mentioned: $\tilde{\Gamma} = \Gamma_x + \mathrm{i}\Gamma_y$
$\mathbf{r}$	Position vector of a point within the Earth relative to the geocenter: $\mathbf{r} = (x, y, z)$
$\mathbf{r}_{\mathbf{B}}$	Position vector of a celestial body B: $\mathbf{r}_{\mathrm{B}} = (x_{\mathrm{B}}, y_{\mathrm{B}}, z_{\mathrm{B}})$ in TRF; relative to the geocenter: $\mathbf{r}_{\mathrm{B}} = (X_{\mathrm{B}}, X_{\mathrm{B}}, X_{\mathrm{B}})$ in CRF
$r_{\mathbf{B}}$	Distance from the geocenter to B
$\tilde{r}_{\mathbf{B}}$	Complex sum of the first two components of $\mathbf{r}_{\mathrm{B}}$ in terrestrial frame: $\tilde{r}_{\mathrm{B}} = x_{\mathrm{B}} + \mathrm{i}y_{\mathrm{B}}$
$W_{\mathbf{B}}(\mathbf{r})$	Gravitational potential of a celestial body B
G	Constant of gravitation
$M_{\mathbf{B}}$	Mass of a celestial body B
$M_{\mathbf{E}}$	Mass of the Earth
$\boldsymbol{\theta}_{\mathbf{B}}$	Colatitude of a celestial body B
$\boldsymbol{\lambda}_{\mathbf{B}}$	Longitude of a celestial body B
$\boldsymbol{\rho}(\mathbf{r})$	Density of matter at $\mathbf{r}$
$\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$	Time-dependent coefficients of xz and yz in degree 2 tesseral tide generating potential (TGP) of degree 2 and order 1 = tesseral potential
$\tilde{\boldsymbol{\phi}}$	Complex sum of the previously mentioned two components: $\tilde{\phi} = \phi_1 + \mathrm{i}\phi_2$
$\tilde{\boldsymbol{\Gamma}}$	Torque as a function of $\tilde{\phi}$ and of the 21-coefficient of the TGP: $\tilde{\Gamma} = -iAe\Omega_0^2\tilde{\phi}$, $\tilde{\Gamma} = \left(\frac{15}{8\pi}\right)^{1/2} iAeF_{21}$
$F_{nm}(\mathbf{r}_{\mathbf{B}}, \boldsymbol{\theta}_{\mathbf{B}}, \boldsymbol{\lambda}_{\mathbf{B}})$	Coefficient of the TGP
$F^*_{nm}(\mathbf{r}_{\mathbf{B}}, \boldsymbol{\theta}_{\mathbf{B}}, \boldsymbol{\lambda}_{\mathbf{B}})$	Coefficient of the TGP
$H^{nm}_{\boldsymbol{\omega}}$	Spectral component amplitude in terms of 'height' of the TGP
$\boldsymbol{\Theta}_{\boldsymbol{\omega}}(t)$	Argument of the tidal component: $\Theta_\omega(t) = \omega t + \mathcal{X}_\omega$
$\boldsymbol{\omega}$	Tidal frequency
$\boldsymbol{\chi}_{\boldsymbol{\omega}}$	Tidal phase
$\boldsymbol{\tau}$	Doodson argument called lunar time: $\tau = 180^\circ + \mathrm{GMST} - s + \lambda$
$\boldsymbol{s}$	Doodson argument, mean tropic longitude of the Moon

h	Doodson argument, mean tropic longitude of the Sun
p	Doodson argument, mean tropic longitude of the Moon's perigee
N'	Doodson argument, mean tropic longitude of the Moon's node
p_s	Doodson argument, mean tropic longitude of the Sun's perigee
GMST	Greenwich hour angle of the mean equinox: $\mathrm{GMST} = \mathrm{GMST}_0 + \Omega_0 t + \cdots = 280.^\circ 4606 + \Omega_0 t + \cdots$
GST	Greenwich hour angle of the true equinox
C_{nm}, S_{nm}	Geopotential coefficients: $C_{n0} = -J_2$
K	Hamiltonian for the computation of rigid Earth nutation
c	Semimajor axis of the orbit of B around the Earth (B could be the apparent Sun); also distance from geocenter of B on a circular motion in the orbital plane (for the simple case considered in the chapter)
p	Mean angular velocity of B around the Earth in uniform motion along a circle of radius c centered at the geocenter: $= 2\pi$ radians per revolution $p^2 = GM_B/c^3$
n_S	Mean angular velocity of the apparent Sun $= 2\pi$ radians per year
F	Normalization factor of $\boldsymbol{\Gamma}$ in the CRF: $F = (3GM_B/2c^3)\,\mathrm{A}e \sin\varepsilon$
$\mathbf{V}$	Vector of which the equatorial part executes an elliptical motion
V_1, V_2	Components of $\mathbf{V}$
A_s	Amplitude of the sine component of (V_1, V_2): $V_1 = A_s \sin(\omega t + \alpha)$
B_c	Amplitude of the cosine component of (V_1, V_2): $V_2 = B_c \cos(\omega t + \alpha)$
ω	Frequency of the motion
α	Phase at time $t = 0$
V^+, V^-	Amplitudes of the circular motions: $V^+ = (A_s - B_c)/2$, $V^- = (A_s - B_c)/2$
V^p	Amplitudes of the circular prograde motions: $V^p = \frac{1}{2}\left(\frac{\omega}{\lvert\omega\rvert} A_s - B_c\right)$
V^r	Amplitudes of the circular retrograde motions: $V^r = -\frac{1}{2}\left(\frac{\omega}{\lvert\omega\rvert} A_s - B_c\right)$
$T_n(\omega_n)$	Nutation transfer function: $T_n(\omega_n) \equiv \frac{\tilde{\eta}(\omega_n)}{\tilde{\eta}\mathrm{Rig}(\omega_n)}$
$T_w(\omega_w)$	Wobble transfer function: $T_w(\omega_w) \equiv \frac{\tilde{m}(\omega_n)}{\tilde{m}\mathrm{Rig}(\omega_w)}$ $T_n(\omega_n) = T_w(\omega_w)$
c_{ij}	Components of the incremental inertia tensor
$\tilde{c}$	Complex sum of two of the previously mentioned components: $\tilde{c} = c_{13} + \mathrm{i}c_{23}$
k	Love number for mass redistribution potential
k_f	Fluid Love number: $k_f = 3(C-A)G/(\Omega_0^2 a^5)$
$\boldsymbol{\Omega}_f$	Incremental core angular velocity vector
m_{fx}, m_{fy}	Incremental core wobble equatorial components
$\tilde{m}_f$	Complex sum of the two previously mentioned components: $\tilde{m}_f = m_{fx} + \mathrm{i}m_{fy}$
$\mathbf{H}_f$	Angular momentum vector of the core
$\tilde{c}_f$	Complex sum of two of the components of the core incremental inertia tensor: $\tilde{c}_f = c_{f13} + ic_{f23}$
k	Compliance for the Earth incremental moment of inertia; concerns the centrifugal and external potential: $k = \frac{k}{k_f} e$
ξ	Compliance for the Earth incremental moment of inertia concerns the core centrifugal potential
γ	Compliance for the core incremental moment of inertia concerns the centrifugal and external potential
β	Compliance for the core incremental moment of inertia concerns the core centrifugal potential
σ_E	CW frequency for a rigid Earth also called Euler frequency: $\sigma_E = e\Omega_0$
$\sigma_{CW}^{elastic}$	CW frequency for an elastic Earth: $\sigma_{CW}^{elastic} = \Omega_0(e-k)$
σ_{CW}	CW frequency for Earth with a fluid core: $\sigma_{CW} = \Omega_0 \frac{A}{A_m}(e-k)$
m_{sx}, m_{sy}	Incremental inner core wobble equatorial components
$\tilde{m}_s$	Complex sum of the two previously mentioned components: $\tilde{m}_f = m_{fx} + \mathrm{i}m_{fy}$
$\tilde{\Theta}_s$	Tilt of the solid inner core: $\dot{\tilde{\Theta}}_s = \tilde{m}_s$
K^{CMB}	Coupling constant at the core–mantle boundary
K^{ICB}	Coupling constant at the inner core boundary
$(x_p - y_p)$	Coordinates of the CIP in the TRF
$\tilde{p}$	Complex sum of the previously mentioned components: $\tilde{p} = x_p - iy_p$ $\tilde{p} = \tilde{p}(\omega_p)\mathrm{e}^{\mathrm{i}(\omega_p t + \chi_p)}$ $\tilde{p} = -(X + iY)\mathrm{e}^{-i\Omega_0(t-t_0)}$ $\tilde{p} = -\tilde{\eta}\mathrm{e}^{-i\Omega_0(t-t_0)}$
ω_p	Frequency of polar motion in TRF: $\omega_p = \omega_n - \Omega_0$

χ_p	Phase of the component of frequency ω_p of $\tilde{p}$
$\tilde{p}(\omega_p)$	Amplitude of the component of frequency ω_p of $\tilde{p}$: $\tilde{p}(\omega_p) = -\tilde{\eta}(\omega_n)$
$\rho_0(r)$	Mean density at radius r
$\mu_0(r)$	Rigidity modulus at radius r: $\mu_0(r) = 0$ in the fluid part
$\kappa_0(r)$	Incompressibility modulus at radius r
$\lambda_0(r)$, $\mu_0(r)$	Lamé parameters: $\lambda_0(r) = \kappa_0(r) - \frac{2}{3}\mu_0(r)$
$\phi_0(r)$	Gravity potential at radius r
$\psi_0(r)$	Centrifugal potential
$\Phi_0(r)$	Initial hydrostatic potential: $\Phi_0(r) = \phi_0(r) + \psi_0(r)$
$\tilde{g}$	Gradient of the gravity potential: $\tilde{g} = \frac{dW_0}{dr}$
g	Gravity: $g = 2\pi G\rho_0 r + \frac{r}{2}\frac{d^2W_0}{dr}$
σ	Frequency expressed in cpsd
$D^l_{mn}(\theta, \lambda)$	Generalized spherical harmonics (GSH)
l	GSH degree: l in $[,+\infty]$
m	GSH order: m in $[-l, +l]$
n	GSH third integer: n in $[-l, +l]$
$\hat{e}_-, \hat{e}_0, \hat{e}_+$	Basis vector used in GSH approach
Φ^{ext}	External gravitational potential
Φ_1^E	Mass redistribution potential: spherical harmonics components $\Phi_{1l}^{Em}(r)$
$\Phi_{1l}^{Em}(r)$	Related to derivative of the mass redistribution potential: $g_{1l}^{Em}(r) = \frac{d\Phi_{1l}^{Em}(r)}{dr} + 4\pi G\rho_0 U_l^m(r)$
s	Displacement vector
$S_l^{m-}, S_l^{m0}, S_l^{m+}$	Components of s
$U_l^m(r)$	Radial component of s: $U_l^m(r) = S_l^{m0}(r)$
$V_l^m(r)$	Spheroidal tangential component of s: $V_l^m(r) = S_l^{m+}(r) + S_l^{m-}(r)$
$W_l^m(r)$	Toroidal tangential component of s: $W_l^m(r) = S_l^{m+}(r) - S_l^{m-}(r)$
X_l^m	Dilatation: $X_l^m = \frac{dU_l^m}{dr} + \frac{1}{r}\left(L_0^l V_l^m + 2U_l^m\right)$
T(r, t)	Incremental stress tensor: components $T_{li}^{mj}(r)$
$P_l^m(r)$	Pressure: $P_l^m(r) = T_{l0}^{m0}(r)$
$Q_l^m(r)$	Spheroidal tangential stress: $Q_l^m(r) = T_{l0}^{m+}(r) - T_{l0}^{m-}(r)$
$R_l^m(r)$	Toroidal tangential stress: $R_l^m(r) = T_{l0}^{m+}(r) - T_{l0}^{m-}(r)$
$\eta_{1l}^{Um}, \eta_{2l}^{Um}, \eta_{3l}^{Um}$, $\eta_{1l}^{Vm}, \eta_{2l}^{Vm}, \eta_{3l}^{Vm}$, $\eta_{1l}^{Wm}, \eta_{2l}^{Wm}, \eta_{3l}^{Wm}$	Intermediate notation used in the scalar equations in the frequency domain and using GSH

Abbreviations

μas	Microarcsecond
CIO	Celestial Intermediate Origin
CIP	Celestial intermediate pole
CMB	Core–mantle boundary
cpsd	Cycle per sidereal day
CRF	Celestial reference frame
CW	Chandler wobble
ECRF	Ecliptic celestial reference frame
FCN	Free core nutation
FICN	Free inner core nutation
FOC	Fluid outer core
GAST	Greenwich Apparent Sidereal Time
GCRF	Geocentric Celestial Reference Frame
GMST	Greenwich Mean Sidereal Time
GSH	Generalized spherical harmonics
GST	Greenwich Sidereal Time
ICB	Inner core boundary
ICRF	International Celestial Reference Frame
ICW	Inner core wobble
ITRF	International Terrestrial Reference Frame
J2000	Event (epoch) at the geocenter at date 2000 January 1.5
LOD	Length of day
mas	Milliarcsecond ($=4.8481 \times 10^{-9}$ rad)
NDFW	Nearly diurnal free wobble (FCN as seen in the TRF)
PFCN	Prograde free core nutation (FICN = PFCN)
SIC	Solid inner core
TGP	Tide-generating potential
TRF	Terrestrial reference frame
VLBI	Very long baseline interferometry

3.10.1 Introduction–Concepts–Overview

3.10.1.1 Manifestations of Variations in Earth Rotation and Orientation

Variations in the orientation in space of an Earth-fixed reference frame (Earth orientation variations) are driven by variations in Earth rotation, that is, in the angular velocity vector of Earth rotation. They are manifested as variations in the direction of Earth-related axes in space (precession and nutation; see Figure 1) as well as those relative to a terrestrial reference frame (TRF) (wobble and polar motion) and also as variations in the angular speed of rotation (or spin rate, for brevity) that translate into variations in the length of day (LOD).

The axes of interest are the figure axis, rotation axis, and the angular momentum axis. The term 'figure axis' is used herein for the axis of maximum moment of inertia of the 'static' Earth, that is, of the Earth with time-dependent deformations caused by tidal and other forces disregarded. (The axis of maximum moment of inertia of the dynamically deforming Earth has indeed been called the figure axis in some of the literature,

but it is of little interest.) The rotation axis is simply the direction of the angular velocity vector (also called the rotation vector). All the axes may be thought of as unit vectors in the directions of the respective axes. Nutations of all three axes are distinct, yet closely related.

With every axis is associated a pole, which is the intersection of the axis with a geocentric sphere of unit radius (in arbitrary units). The motion of the axis is faithfully reflected in the motion of the associated pole and also in the motion of the plane perpendicular to the axis (the equator of that axis), so the nutation, for instance, may and often is referred to as the nutation of the associated pole or equator.

Precession of any axis is the secular (smoothly varying) part of the motion of the axis in space, that is, relative to the directions of 'fixed' stars. (Closest to the ideal of fixed stars are the quasars, the most distant objects in space.) It is the 'mean' motion of the axis in space. Nutation is the oscillatory part of the motion of the axis, with a large multiplicity of periods; it causes the instantaneous position of the axis to move around its mean position along an oscillatory path (see Figures 1–3).

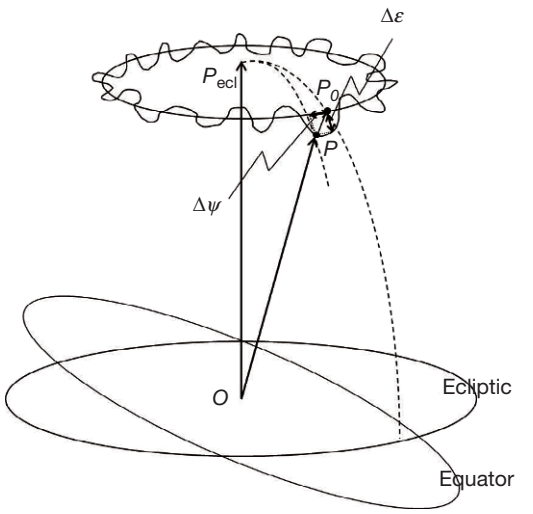


Figure 1 Precession and nutation of an Earth-related axis. Precession carries the pole of the axis at a uniform rate in an anticlockwise sense along a circle on the surface of the celestial sphere, centered on the normal to the ecliptic plane. Nutation consists of small deviations from this uniform motion, both over the precessional path (nutation in longitude, $\Delta\psi$) and perpendicular to it, on the celestial sphere (nutation in obliquity, $\Delta\varepsilon$). The resulting path of the pole appears wiggly, as in the figure.

The term 'wobble' is used herein in a very broad sense, for any periodic or quasiperiodic motion of the instantaneous rotation axis with respect to the figure of the Earth, irrespective of the frequency or the physical origin of the motion. The essence of wobble is the temporally varying offset (angular separation) of the direction of the rotation axis from that of the figure axis. Wobble is characterized by the variation of the equatorial components of the rotation axis in a terrestrial frame of reference. The Earth is made up broadly of three regions, the mantle and the two core regions within it, the fluid outer core (FOC), and the solid inner core (SIC). The term 'solid Earth,' as generally used, includes all three regions while excluding the fluid layers at the surface, namely, the oceans and the atmosphere. The three regions of the solid Earth have different rotation vectors, though their motions are mutually coupled. Therefore, each region has its own wobble. The term 'wobble,' when used without specific mention of the region, is intended to be that of the mantle.

The term 'polar motion' has been in use to refer generally to any motion of the Earth's pole of rotation relative to the Earth and, in particular, to the wobble as well as for secular polar motion. The latter is a steady drift of the mean pole of rotation (with the periodic wobbles averaged out) relative to the Earth; it may be viewed as the present-day trend of the polar wander, which has been taking place over geologic timescales. In recent usage, polar motion has taken on a technical meaning, to be explained later on (see Sections 3.10.2.5.1 and 3.10.12.4), as the motion of a conceptual pole called the celestial intermediate pole (CIP) with frequencies outside a specified range. A different aspect of the variations in Earth rotation is the variation in its spin rate or, its equivalent, the LOD variation. The mean rate of Earth rotation is one cycle per sidereal day (cpsd), equivalent to an angular speed $\Omega_0 = 2\pi$ radians per sidereal day $= 7.292115 \times 10^{-5}$ rad s^{-1}. (The sidereal day is the period of rotation of the Earth relative to the most distant

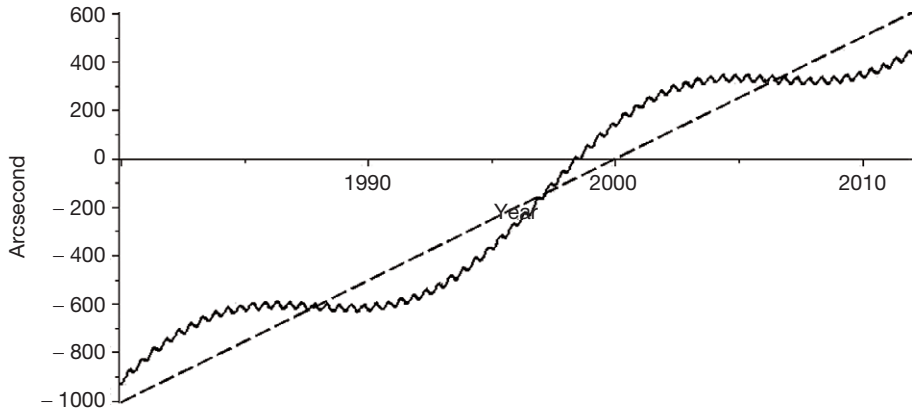


Figure 2 Precession and nutation of the axis in longitude. The dashed straight line represents the precessional motion, and the deviation of the curve from the straight line represents nutation. To make this deviation clearly visible, the angle of nutation has been scaled up by a factor of ten relative to the precession.

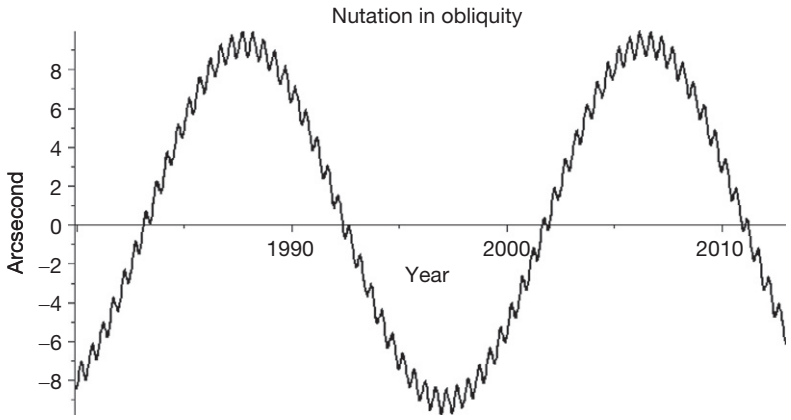


Figure 3 Nutation of the axis in obliquity.

'fixed' stars. It is $T_d/(T_d+1)$ times the solar day of 24 h, where T_d, equal to 365.2422, is number of solar days in the period of the relative orbital motion of the Earth and the Sun. The solar day is the period of rotation of the Earth relative to direction toward the Sun.) The small deviations from the uniform rate of rotation give rise to LOD variations.

3.10.1.2 Free Rotational Modes

The variations in Earth rotation/orientation are made up of both free motions and forced motions of the various axes, apart from variations in the spin rate.

The free motions, which are not driven by any external torque, conserve the Earth's total angular momentum. The angular momentum is a product of the moment of inertia (which, in general, is a tensor) and the angular velocity, and its conservation requires that if one changes with time, the other must undergo compensating changes. This happens whenever the rotation (angular velocity) vector of the mantle or other region of the Earth is not along a principal axis, causing the inertia tensor (in a space-fixed frame) to vary as the rotation progresses; this in turn, causes a compensatory variation of the rotation vector. The free nutation/wobble normal modes are the manifestations of this phenomenon. The three-layer Earth has four such modes: the Chandler wobble (CW) with a period of about 430 days; the free core nutation (FCN) with a period of about −430 days in the celestial frame, along with its associated wobble (the nearly diurnal free wobble or NDFW), which has nearly diurnal frequency in the terrestrial frame as its name implies; the less prominent free inner core nutation (FICN, also called the prograde free core nutation or PFCN), which has a rather long period, apparently around 1000 days, in the celestial frame; and the inner core wobble (ICW) with a still longer period in the terrestrial frame. See Annex **A** for further explanations. They are excited by various geophysical processes. They have been found, observationally, to have variable amplitudes and are not amenable to theoretical prediction. Since the frequencies of the FCN and FICN/PFCN modes lie in the midst of the low-frequency band wherein the most important nutation components lie, the resonances associated with these modes play an important role in the nutational responses of the Earth. Though the CW frequency lies well outside this band when viewed in the celestial frame, its influence on nutation is not completely negligible unlike that of the ICW.

3.10.1.3 Forced Motions

The forced nutation, wobble, and precessional motions of the axes are excited almost wholly by the gravitational pulls of the Sun, Moon, and planets on the nonspherical (nearly ellipsoidal) Earth and are different aspects of the gyroscopic response of the rotating Earth to the gravitational torque. The deformations produced by the direct action of the spatially varying gravitational force (as distinct from the gravitational torque) on the solid Earth, as well as by the loading effect of ocean tides and the gravitational attraction of the solid Earth by the redistributed ocean mass, play an important role in rotation variations; so does the deforming action of the centrifugal perturbation caused by the rotation variations themselves. Torques exerted by atmospheric pressure tides (arising primarily from the varying thermal input from solar radiation) also make minor contributions to variations in the directions of the axes. The forced motions of the axes due to the gravitational torques on the solid Earth are accurately predictable, given a sufficiently accurate knowledge of the relevant Earth parameters and of the orbital motions of the various celestial bodies.

The nutational motions consist of numerous periodic components with a discrete spectrum of frequencies determined by the spectrum of the gravitational potential, at the Earth, of the solar system bodies. The frequencies of spectral components with nonnegligible amplitudes extend from large negative values to large positive values for the nutations of any of the axes mentioned in Section 3.10.1.2. (By 'large,' we mean several times the frequency of the diurnal Earth rotation.) Positive frequencies correspond to prograde motions, which are in the same sense as that of the diurnal Earth rotation, while negative frequencies are for retrograde motions (in the opposite sense).

Figure 4 shows the observations of LOD variations and polar motion.

The dominant causes of polar motion (other than the tidally excited wobbles) are the atmosphere and the ocean, at timescales of a few days to a few years in the terrestrial frame (see Chapter 3.09). These geophysical fluids are also primarily responsible for the LOD variations at timescales of a few days to a few years in the terrestrial frame, while the interaction

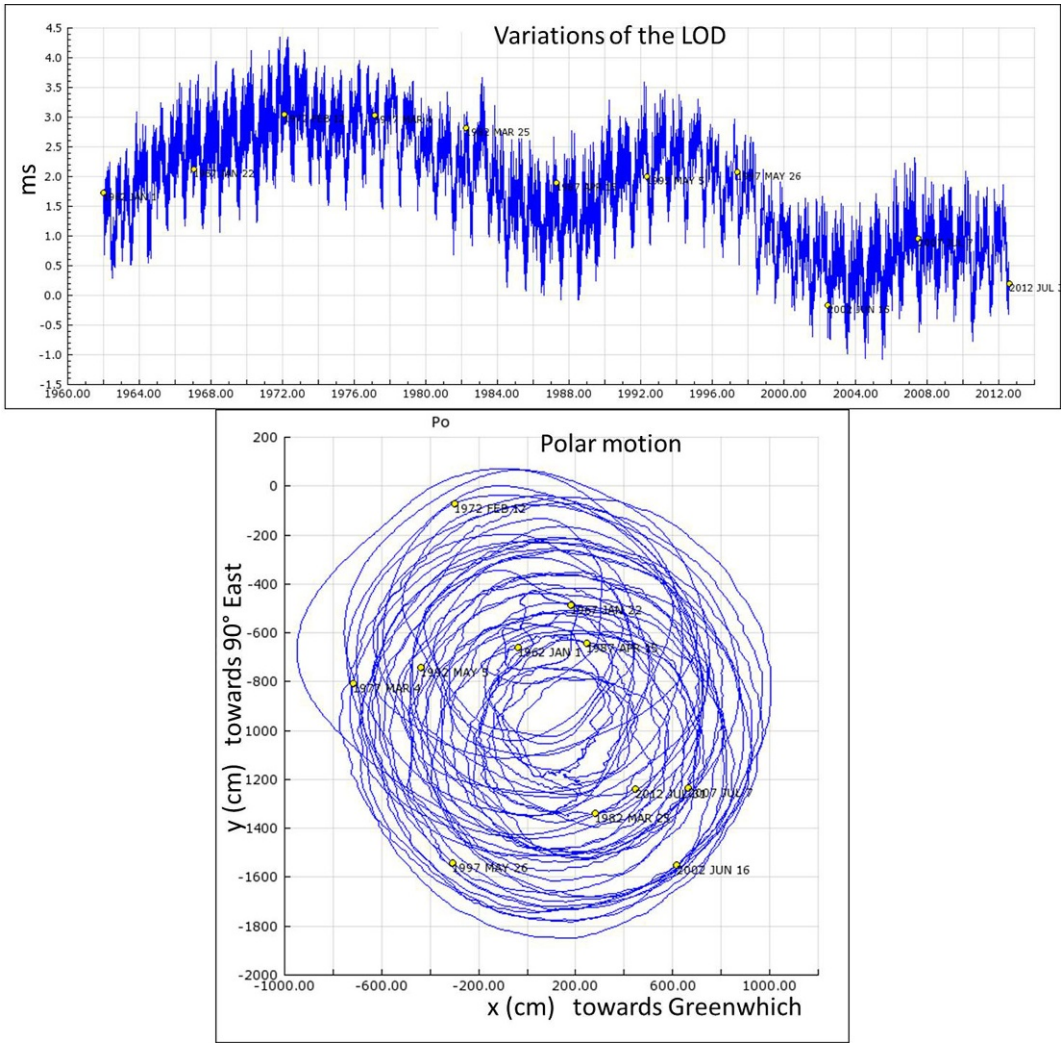


Figure 4 Observation of length-of-day variations and polar motion. Figure built from the IERS website.

of the magnetic field in the fluid core with the mantle is invoked to explain the variations at decadal timescales (see Chapter 3.09). There is a discrete spectrum in the LOD variations, arising primarily from the changes in the axial moment of inertia resulting from the deforming action of the tidal potential and to a very much smaller extent by tidal torques: the axial component of the tidal torque depends on the difference between the two principal moments of inertia in the equatorial plane (A and B), which is less than 1% of the difference between the polar (C) and equatorial moments of inertia.

3.10.1.4 Earth's Response to Gravitational Torques

The extent of the Earth's response to gravitational torques is determined by the values of a number of Earth parameters, which represent various aspects of the structure and properties of the Earth's interior: the matter distribution within the Earth, the existence and mutual couplings of the FOC and SIC regions, the couplings of these regions to the mantle, and the rheological and other properties of the Earth's interior. By far, the most important of the Earth's parameters is the dynamical ellipticity, which is the fractional difference between the moments of the Earth's inertia about the polar and equatorial axes; it is a reflection of the Earth's 'equatorial bulge.' The response to the external forcing is modified by the tidal and other excitations of the fluid layers at the surface, namely, the oceans and the atmosphere.

3.10.1.5 Low-Frequency Nutations

The spectrum of the nutational motion is discrete and is determined by the spectra of the orbital motions of the solar system bodies. The major spectral components of the nutation have frequencies within the low-frequency band (between −0.5 and 0.5 cpsd); many of these components have amplitudes that are larger than those of the spectral components in high-frequency bands (frequencies exceeding 0.5 cpsd) by five orders of magnitude. The low-frequency forced nutations arise from the action of external gravitational potentials on the axially symmetrical part of the matter distribution within the Earth. The distribution is, in fact, both axially symmetrical and

ellipsoidal to a very high degree of approximation. The principal moments of inertia are then A, B, and C, with $C > B > A$ and $B \approx A$, and the dominant role in Earth rotation variations (other than those in the LOD) is played by the dynamical ellipticity defined either by $H_d = (C-A)/C$ or by $e = (C-A)/A \approx 1/304.5$, the former parameter being favored by astronomers and the latter by geophysicists. There exists a residual (nonellipsoidal) axially symmetrical part of the density distribution; it too gives rise to low-frequency nutations but with amplitudes much lower than those of the dominant nutations due to e. These are closely followed by the semidiurnal nutations (frequencies between 1.5 and 2.5 cpsd) arising from the triaxiality of the Earth, that is, the small fractional difference $(B-A)/A$ between the two equatorial moments of inertia; its value happens to be about $2e^2$, of the second order in the dynamical ellipticity. The high-frequency nutations, for which the nonaxially symmetrical part of the density distribution is responsible, have even smaller amplitudes as already stated.

In view of the preceding text, most theoretical formulations of Earth rotation variations are, justifiably, based on axially symmetrical ellipsoidal Earth models characterized by the dynamical ellipticity e and are therefore concerned with low-frequency nutations only.

3.10.1.6 Gravitational Perturbations and Displacement Fields

Different elements of matter in the Earth are subjected to differing gravitational accelerations by the Moon (or Sun or other body), because they are at different distances from the external body. Therefore, they undergo position-dependent displacements from their unperturbed positions (i.e., positions in the absence of the gravitational forces). In the case of a deformable Earth model, the displacement field $\mathbf{s}(\mathbf{r})$ over the whole Earth can be analyzed into a rotational part wherein the relative distances between different points remain undisturbed, and a deformational part. One approach to the theory of the gravitational perturbations of the Earth seeks to solve the equation of motion governing the field $\mathbf{s}(\mathbf{r})$, given a detailed model of the matter density and elastic constants as functions of position in the Earth's interior, to determine thereby both the deformations (the so-called solid Earth tides) and the rotational motions (nutations and precession) produced by the perturbing potential. Treatments employing the displacement field approach use only axially symmetrical ellipsoidal Earth models (see Dehant, 1987a,b; Smith, 1974; Wahr, 1981). Spherically symmetrical models are constructed from information provided by seismic data; the axially symmetrical ellipsoidal structure is then derived from such models using a theory (hydrostatic equilibrium) that determines the flattening of the constant density surfaces that is caused by the centrifugal effects of the Earth's mean rotation at a constant rate around the polar axis. Among the most widely used models is PREM of Dziewonski and Anderson (1981), modified to the ellipsoidal structure based on hydrostatic equilibrium theory. Further developments in the direction of incorporating nonhydrostatic equilibrium and triaxiality inside the Earth has only been undertaken very recently and with the objective of application to other planets or moons of our solar system than the Earth (Trinh et al., 2011).

From this spherically symmetrical Earth model, one has to go over to one that is ellipsoidal in structure when studying Earth rotation variations. The hypothesis of hydrostatic equilibrium is invoked to go over to an axially symmetrical ellipsoidal model: ellipticity of the Earth is an essential ingredient in the theory of rotation variations, as we have repeatedly seen.

3.10.1.7 Gravitational Torque on the Earth

A simpler approach focuses directly on the rotation variations by employing the equation of motion governing rotational motion, namely, the angular momentum balance equation (torque equation). The gravitational torque on the Earth is dependent solely on the matter distribution within the Earth and on the positions of the gravitating bodies relative to the TRF. Information on the motions of the solar system bodies is derived from the ephemerides. As for the matter distribution, one could rely on detailed models of the spherically symmetrical density profiles within the Earth constructed from seismic studies, which are converted to axially symmetrical ellipsoidal models constructed in the manner indicated in the last section. The torque on an idealized model with the previous symmetry involves only one Earth parameter, namely, the dynamical ellipticity, which represents the fractional difference between the moments of inertia about polar and equatorial axes. This parameter is sufficient to account for all but a very small part of the nutational motion. But when a fuller account is needed, one has to consider other features of the Earth's structure, which are represented by the higher-degree geopotential coefficients, on which other parts of the torque depend. These coefficients are estimated from space geodetic observations of the orbits of artificial Earth satellites.

3.10.1.8 Earth Response to the Torque

When considering the rotation variations of a wholly solid model Earth, its angular momentum can be expressed as a product of its inertia tensor and angular velocity vector. The torque equation in the rotating terrestrial frame then becomes an equation for the angular velocity components; the wobble and spin rate variations are related directly to these components and may be obtained by solving this equation. But when the existence of a fluid core region and the freedom it has to rotate with a different angular velocity from that of the mantle are taken into account, the torque equation for the whole Earth involves the angular velocities of both the mantle and the fluid core, and so, one needs to consider also the torque equation for the core region alone. Couplings of the core to the mantle enter into the latter equation and influence the rotation variations of the mantle that are observable. Examples of such couplings are the so-called inertial coupling due to the fluid pressure on the ellipsoidally shaped core–mantle boundary (CMB) and the electromagnetic (EM) coupling and possibly also viscous coupling at the CMB. The differential wobble motions of the mantle and the fluid core, on the one hand, and the presence of magnetic fields crossing over from the fluid core to the mantle side, on the other hand, cause induction of currents in the conducting layers at the bottom of the mantle; and the interaction of the magnetic field with the induced currents all

over the conducting mantle layer is responsible for the EM coupling. Simultaneous solution of the equations for the whole Earth and for the core yields the wobbles of both the mantle and the core. Inclusion of the inner core into the theory brings in a pair of additional equations, one governing its angular velocity variations and the other, its orientation relative to the mantle.

3.10.1.9 Role of Deformations in Earth Rotation

The inertia tensor of each of the regions (or layers) is not just that of the unperturbed Earth: It includes the perturbations associated with deformations produced directly by the gravitational potentials of the external bodies as well as by the centrifugal potentials due to the variations in the angular velocity vectors of all three layers. The perturbations of each of the inertia tensors are made up, therefore, of parts that are proportional to the strength of the perturbing potential and to the wobbles of the three different regions. The proportionality constants are deformability parameters (also called compliances), which are intrinsic properties of the Earth model; they may be computed by integration of the equations governing the deformations of the Earth, given a detailed model of the matter density and elastic constants as functions of position in the Earth's interior. While the compliance parameters are real for an elastic Earth, the anelastic behavior of the mantle results in small complex increments to their elastic-Earth values. Loading of the crust by the ocean tides raised by the gravitational attraction of the external bodies causes further deformation, which again increments the inertia tensors and thereby influences the rotation variations. The increments due to ocean tidal effects, including the effect of ocean tidal currents, are nontrivial and have been calculated by different methods. One method that has been recently employed represents the deformational effect of ocean tides in terms of effective increments to the compliance parameters, which are not only complex but also frequency-dependent.

3.10.1.10 Interplay of Various Phenomena

In order to get results of high accuracy for Earth rotation variations, it is also necessary to take into account the mutual influences of a variety of phenomena. The tidal deformations of the Earth, including the contribution from the centrifugal perturbation caused by Earth rotation variations, as well as from ocean loading, affect the rotations of all three regions through the changes in their inertia tensors; the deformations also perturb the gravitational potential of the Earth itself, with a reciprocal effect on the deformations themselves; they affect the ocean tides through movements of the ocean floor and through the effect of the perturbation of the Earth's gravitational potential on the ocean surface; and the consequent modification of the ocean tidal loading and tidal currents influences both the deformations and the rotation variations. In brief, there is an interplay of the various phenomena, which is a factor in determining the effects of each one of them, and it is necessary to take account of this interplay while seeking to obtain accurate results for any of these effects. A chart summarizing all the interactions between the phenomena involved is presented in Figure 5.

3.10.1.11 Precision of Observations: Challenge to Theory

The precision of recent nutation estimates has surpassed 0.2 milliarcseconds (mas), and that of the estimates of individual spectral components is of the order of 10 microarcseconds (μas) for many of the more prominent components except for those with periods exceeding 1 year, for which the uncertainties are

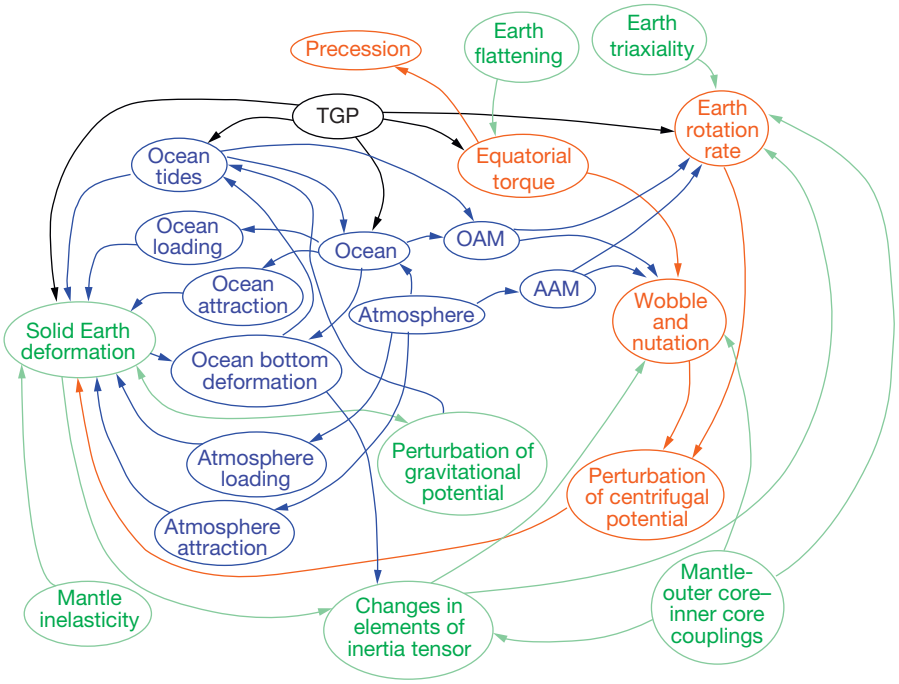


Figure 5 Chart representing the interactions and physical processes involved in Earth orientation variations and in Earth deformations.

higher ($1\ \text{mas} = 4.848 \times 10^{-9}$ rad, $1\ \mu\text{as} = 4.848 \times 10^{-12}$ rad). To match this level of precision, all effects that make contributions at the μas level to nutations have to be identified and taken into account. One such effect that had been ignored till very recently is the torque produced by the action of the gravitational potential of solar system bodies on the perturbations in the matter distribution within the Earth due to the action of the potential itself. Recent work has shown that the small contributions to nutation and precession arising from such second-order torque terms are not negligible in this context. General relativistic effects (geodesic precession and nutation) and effects of atmospheric pressure are also significant.

3.10.2 Earth Orientation/Rotation Variables: Reference Frames

As a preliminary observation, we note that the direction of any of the coordinate axes or other axes like the Earth's rotation axis or the angular momentum axis may be conveniently represented by its point of intersection with a geocentric unit sphere called the celestial sphere; the point is called the pole of the axis. Variations of the coordinates of the pole reflect the motions of the axis. The definition of the variables representing the motions of a chosen axis (or its pole) calls for the use of a suitable reference frame. For motions in space (precession and nutation), a celestial reference frame (CRF) has to be used, while wobble and polar motion are to be referred to a terrestrial frame. The celestial and terrestrial frames considered here are all geocentric: the coordinate axes originate from the geocenter.

There are two types of CRFs in use.

3.10.2.1 Ecliptic Celestial Reference Frame

It has the ecliptic as the principal plane (the plane containing the first two Cartesian coordinate axes) to which the third axis is perpendicular. The ecliptic plane may be thought of as the plane of the Earth's orbit around the Sun (or more correctly, as the plane of the orbit of the center of mass of the Earth–Moon system). The circle along which this plane intersects the celestial sphere provides a useful representation of the ecliptic plane and may also be referred to as the ecliptic, and the intersection point of the normal to the ecliptic with the celestial sphere is called the pole of the ecliptic. Similarly, the intersection of the equatorial plane (perpendicular to the axis of interest) with the celestial sphere will be called the equator, and the intersection of the axis itself with the celestial sphere is known as the pole of the equator. The intersections of the equator and the ecliptic are the equinoxes. The equinox at which the Sun crosses over from south of the equator to the north in its annual journey along the ecliptic (as perceived from the Earth) is the vernal (spring) equinox and the other is the autumn equinox. The first axis of the ecliptic reference frame is chosen to be along the nodal line (line of intersection) of the ecliptic and equatorial planes and is taken in the direction of the vernal equinox.

3.10.2.1.1 Precession

The angle between the equatorial and ecliptic planes (or between their polar axes) is the obliquity ε. A steady rotational motion of the equatorial plane or its polar axis about the first axis of the ecliptic frame constitutes the precession in obliquity, which is extremely small and is often ignored; the additional oscillatory (multiply periodic) rotation about the same axis is the nutation in obliquity, $\Delta\varepsilon$. The precession in longitude is a steady rotational motion of the Earth-related axis around the normal to the ecliptic plane, keeping an angle ε_0 to this normal, where $\varepsilon_0 = \varepsilon - \Delta\varepsilon$. Precession takes the pole of this axis over an arc of a circle on the celestial sphere, which is of radius $\sin\varepsilon_0$ and is centered on the normal. The angular size of this arc as seen from the center of the circle is denoted by $\psi_A(t)$; its linear dimension is evidently $\psi_A(t)\sin\varepsilon_0$. The nutation in longitude is an additional oscillatory rotation $\Delta\psi$ about the same normal (over the same arc). The direction of the precessional motion of the Earth's pole is defined to be the direction of increasing longitude; it is opposite to that of the Earth–Moon system around the normal to the ecliptic.

At the present rate of precession, which is very close to 50 arcseconds per year, the pole of the Earth's axis would take about 25 770 years to complete a full circuit around the pole of the ecliptic.

In defining an ecliptic-based frame, one has to contend with a complication: The ecliptic itself has a very slow rotation in space around a vector lying in the ecliptic. One may choose a reference frame based on the 'ecliptic of epoch,' that is, the ecliptic at a fixed epoch (instant of time). The standard choice of epoch is J2000, which is at 12 h Universal time on 1 January, 2000. Alternatively, one may choose a frame based on the ecliptic plane of the time of interest (ecliptic of date). The nutation and precession referred to the fixed ecliptic represent the actual motion of the nutating axis in space. The precession in this sense is only in longitude; it is often called the lunisolar precession because it is caused almost entirely by torques exerted by the Moon and Sun. But the classical definition of nutation and precession is with reference to the moving ecliptic; the precession in longitude thus defined includes a small part arising from the motion of the ecliptic and is called the general precession. There is also a small precession in obliquity arising from the ecliptic motion.

3.10.2.1.2 Nutation

The classical nutation variables $\Delta(t)$ and $\Delta\varepsilon(t)$ representing the oscillatory part of the motions in longitude and obliquity were defined, strictly speaking, relative to the moving ecliptic. The purely periodic components of both $\Delta\psi$ and $\Delta\varepsilon$ remain unaltered even if the fixed ecliptic is used, though the so-called Poisson terms do change (see Bretagnon et al., 1997). Poisson terms are products of t or powers of t multiplied by sine or cosine factors.

A representation of the nutational motion is provided in Figure 6.

$\Delta\psi$ goes up to about ± 19 arcseconds and $\Delta\varepsilon$ up to about ± 10 arcseconds.

3.10.2.2 Geocentric Celestial Reference Frame

This is a frame having the mean equatorial plane of J2000 as its principal plane and having no rotation in space. The qualification 'mean' in this context means that the nutational part of the displacement of the figure axis/equator is excluded while fixing the position of the equator at this epoch. The first axis

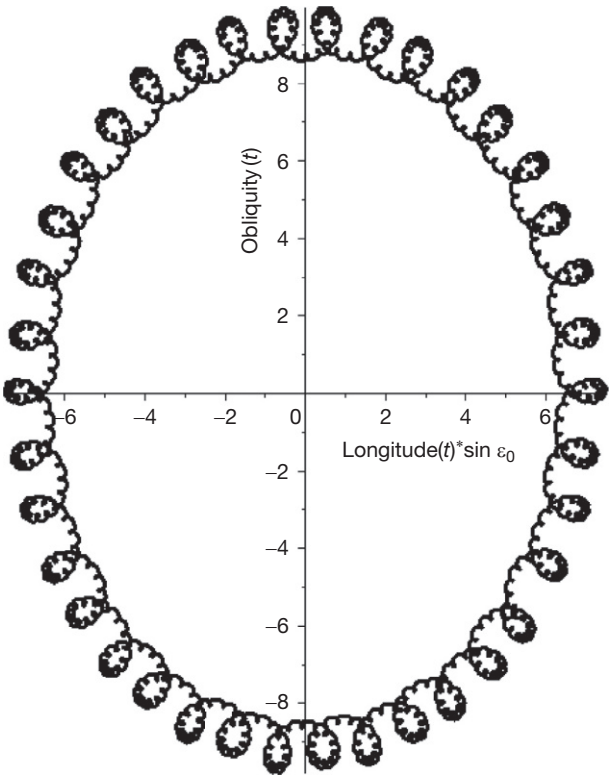


Figure 6 The trace, on the celestial sphere, of the nutational part alone (excluding precession) of the motion of the pole of an Earth axis. The position of the pole at J2000 is indicated. The main feature, the overall elliptical shape, is a reflection of the dominant 18.6-year nutation, and the loops over the ellipse represent primarily the semiannual nutation. The motion of the pole around the ellipse is in the clockwise sense.

(X-axis) of the frame is in the direction of the vernal equinox. This is the direction of increasing longitude referred to the ecliptic of J2000, and the Y-axis of the Geocentric Celestial Reference Frame (GCRF) is in the direction of increasing obliquity.

In the GCRF, the displacement of the pole of the axis of interest due to the precession and nutation is represented by the coordinates $X(t)$, $Y(t)$. For times that are not too far from J2000, one may take

$$\begin{aligned} X(t) &= [\psi_A(t) + \Delta\psi(t)] \sin \varepsilon_0 \\ Y(t) &= \varepsilon_A(t) + \Delta\varepsilon(t) \end{aligned} \qquad [1]$$

where $\varepsilon_A(t)$ is the mean obliquity ε_0 at J2000 plus the precession in obliquity. But when the time difference from J2000 gets large, corrections have to be applied to these lowest-order expressions. The rate of precession in obliquity, $\dot{\varepsilon}_A(t)$, is only about 0.25 milliarcseconds per year (mas year^{-1}), which is extremely small compared with $\dot{\psi}_A(t)$, which is about 50 400 mas year^{-1}.

3.10.2.3 International Terrestrial Reference Frame

A terrestrial frame, by definition, is corotating with the Earth. The International Terrestrial Reference Frame (ITRF) is realized by a set of three mutually orthogonal 'Earth-fixed' axes, with its origin at the center of mass of the unperturbed Earth. The positions of the axes are determined by the coordinates inferred for an adequate number of observation stations on the *unperturbed* Earth from their positions on the real Earth as determined by appropriate observations (e.g., from a network of very long baseline interferometry (VLBI) stations); one has to remove, from the station position data thus obtained, the time-dependent displacements suffered by the station through the effects of solid Earth tides, tectonic plate motions, etc., through the use of suitable models. The principal plane of this reference frame is the plane of the true equator, perpendicular to the instantaneous direction of the figure axis, which is the third axis of the frame; its first axis (x-axis) lies on the Greenwich meridian, and the y-axis direction, orthogonal to the other two, is such as to make the frame a right-handed one. When we refer to a terrestrial frame, it is to be understood, by default, to be the ITRF unless otherwise specified.

3.10.2.4 Principal Axis Frame

This is a terrestrial frame that differs from the ITRF only in having the two equatorial coordinate axes along the two principal axes of inertia in the equatorial plane. Its x-axis lies about 15° west of the Greenwich meridian. In setting up the torque equations governing the rotational motion of the Earth, it is convenient and customary to make use of this reference frame.

The wobble variables are defined, whether in the ITRF or in the principal axis frame, as

$$m_x = \Omega_x/\Omega_0 \quad \text{and} \quad m_y = \Omega_y/\Omega_0 \qquad [2]$$

where Ω_x and Ω_y are the first two components of the angular velocity vector $\boldsymbol{\Omega}$ of the Earth (or more correctly, of the mantle) and Ω_0 is the mean angular velocity. Wobble is an inescapable accompaniment of nutation and vice versa. Wobble is intimately related to the nutation of the figure axis, as we shall see, and plays an essential role in various approaches to the theoretical treatment of nutation.

3.10.2.5 The CIP and CIO

The most direct and most precise information about the varying orientation of the TRF in space is obtained from VLBI. This technique uses observations of the time delays between the arrival times of radio signals from the most distant (and hence the most 'space-fixed') sources of radio waves in space at different members of an array of radio antennas spread over the globe to obtain precise estimates of the orientation of the terrestrial frame relative to a celestial frame defined by such sources. The observations are made at intervals of several days (typically five or seven) and have been carried out for over two decades; they make it possible to estimate with very high precision the variations in orientation, that is, precession, obliquity rate, nutations of the figure axis, and rotation rate variations, having periods exceeding a few days. Similarly, it is the low-frequency variations in direction of the rotation axis in the TRF that are susceptible to estimation with high precision. The conventions adopted by the International Astronomical Union (IAU) conform to this reality.

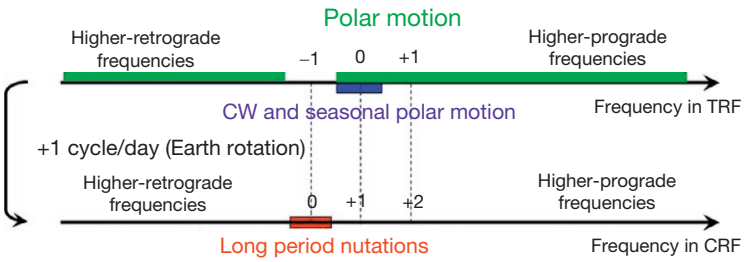


Figure 7 Ranges of polar motion frequencies (marked in green) and of nutation–precession frequencies (marked in red). Precession is induced by the zero-frequency component of the gravitational spectrum in the celestial frame.

3.10.2.5.1 Nutation and polar motion in the conventions: CIP

A pole called the CIP, intermediate between the poles of the terrestrial and CRFs, is introduced such that its nutation, that is, its motion relative to the space-fixed frame (GCRF), has only low-frequency spectral components (between −0.5 and 0.5 cpsd; see Figure 7). It is this motion of the CIP that is defined in the conventions as nutation. This low-frequency band in space maps into the 'retrograde diurnal' band in the terrestrial frame, with frequencies between −1.5 and −0.5 cpsd, because of the prograde rotation of this frame in space at the rate of 1 cpsd. Motions of the CIP in the terrestrial frame with frequencies outside this band are defined, according to the conventions, as polar motion. The low-frequency part of this band, which includes the CW and the tidally induced wobbles, dominates the polar motion. However, the amplitude of the polar motion is quite different from that of the wobble to which it is related, as will be shown in Section 3.10.7 (see also Capitaine, 1990, 2000; Capitaine et al., 2000, 2003).

3.10.2.5.2 LOD variations in the conventions: CIO

Variation in the spin rate around the z-axis, represented by m_z, and the consequent variation ΔLOD in the LOD are related to the z component Ω_z of $\mathbf{\Omega}$:

$$\begin{aligned} m_z &= (\Omega_z/\Omega_0) - 1 \\ \Delta\text{LOD} &= -(\text{LOD})m_z \end{aligned} \qquad [3]$$

In the scheme of analysis of VLBI data, variations in the angular velocity component in the direction of the CIP are estimated rather than those in the z component; these observations are performed at the mas level. Computation of the LOD variations from the spin rate variations estimated in this manner is in accord with the current conventions. The spin rate, in this context, is the rate of change of the angle of rotation around the CIP axis, measured from a point on the equator of the CIP, which is so chosen as to have no rotation about this axis. This point is called the Conventional International Origin (CIO; see Capitaine, 1986, 1990, 2000; Capitaine and Guinot, 1988; Capitaine et al., 1986, 2000, 2003; Guinot, 1979).

3.10.3 Equations of Rotational Motion

The most direct approach to the theoretical treatment of Earth rotation variations is based on the torque equations, which are statements of angular momentum conservation.

3.10.3.1 Equation of Motion in a Celestial Frame

In an inertial (space-fixed) reference frame, the torque equation is simply

$$\frac{\mathrm{d}\mathbf{H}}{\mathrm{d}t} = \mathbf{\Gamma} \qquad [4]$$

So we have, trivially,

$$\mathbf{H} = \int \mathbf{\Gamma}\,\mathrm{d}t \qquad [5]$$

This result is independent of the structure of the body that is subjected to the torque.

The Earth's rotation axis being very close to the normal to the Earth's mean equator and hence to the Z-axis of the reference frame, the components H_X, H_Y are very small compared with H_Z, which is closely approximated by $C\Omega_0$; so $|\mathbf{H}| \approx C\Omega_0$. The X and Y components of the *unit vector* in the direction of $\mathbf{H}$, which characterize the precession and nutation *of the angular momentum axis*, are then

$$\begin{aligned} X_H &= \frac{H_X}{C\Omega_0} = \frac{\int \Gamma_X\,\mathrm{d}t}{C\Omega_0} \\ Y_H &= \frac{H_Y}{C\Omega_0} = \frac{\int \Gamma_Y\,\mathrm{d}t}{C\Omega_0} \end{aligned} \qquad [6]$$

where the subscript H identifies quantities pertaining to the angular momentum axis. The precession and nutation in longitude, $\psi_{A,H}$ and $\Delta\psi_H$, and the nutation $\Delta\epsilon_H$ in obliquity are related to X_H and Y_H through equations of the same form as eqn [1]:

$$X_H = (\psi_{A,H} + \Delta\psi_H)\sin\epsilon, \quad Y_H = \Delta\epsilon_H \qquad [7]$$

Explicit evaluation of the nutation quantities has to wait till the torque acting on the Earth and its time dependence are determined.

3.10.3.2 Equation of Motion in a Terrestrial Frame

Consider now the description of the Earth's rotational motion as seen from an Earth-fixed reference frame. The frame itself is rotating in space, and we denote its angular velocity vector by $\mathbf{\Omega}$. The torque equation governing the variation of the angular momentum vector $\mathbf{H}$ in this reference frame is

$$\frac{d\mathbf{H}}{dt} + \boldsymbol{\Omega} \times \mathbf{H} = \boldsymbol{\Gamma} \quad [8]$$

where the vectors $\mathbf{H}$ and $\boldsymbol{\Gamma}$ are now referred to the terrestrial frame. On writing the equation out in terms of the Cartesian components, we get

$$\begin{aligned} \dot{H}_x + \Omega_y H_z - \Omega_x H_y &= \Gamma_x \\ \dot{H}_y + \Omega_z H_x - \Omega_x H_z &= \Gamma_y \\ \dot{H}_z + \Omega_x H_y - \Omega_y H_x &= \Gamma_z \end{aligned} \quad [9]$$

where $\dot{H}_x = dH_x/dt$, and so on.

The aforementioned equations, unlike the corresponding ones in an inertial frame, involve not just $\mathbf{H}$ but $\boldsymbol{\Omega}$ too. Therefore, it is necessary, in order to solve these equations, to make use of the relationship of the former to the latter. If the Earth were wholly solid, the relation would be $\mathbf{H} = [\mathbf{C}] \cdot \boldsymbol{\Omega}$, where $[\mathbf{C}]$ is the Earth's inertia tensor. This is no longer true if core regions are present. While the mantle rotates with angular velocity $\boldsymbol{\Omega}$, the other regions rotate, in general, with angular velocities differing from $\boldsymbol{\Omega}$; these differences would result in additional contributions to $\mathbf{H}$ from the core regions. Furthermore, the inertia tensors of the different regions will have to include contributions from tidal deformations of the respective regions when the deformability of the Earth is to be taken into account. Thus, it is necessary to use as inputs various aspects of the Earth's structure and properties before the torque equation in the terrestrial frame can be formulated, and additional equations that govern the variation of the angular velocities of the core regions will have to be considered simultaneously. We consider first the simplest case of a rigid Earth.

3.10.3.2.1 Euler equations for a rigid body

For a rigid body, equations in eqn [9] reduce to the Euler equations when coordinate axes parallel to the principal axes of the body are chosen. With such a choice,

$$H_x = A\Omega_x, \quad H_y = B\Omega_y, \quad H_z = C\Omega_z \quad [10]$$

where A, B, C stand for the principal moments of inertia in increasing order of magnitude, as usual. Substituting into eqn [9], we obtain

$$\begin{aligned} A\dot{\Omega}_x + (C-B)\Omega_y\Omega_x &= \Gamma_x \\ B\dot{\Omega}_y + (A-C)\Omega_z\Omega_x &= \Gamma_y \\ C\dot{\Omega}_z + (B-A)\Omega_x\Omega_y &= \Gamma_z \end{aligned} \quad [11]$$

These are the celebrated Euler equations.

For the rotation variations of the Earth, the fractional variations in the angular velocity, namely, m_x, m_y, m_z, are far smaller than unity. (For the real Earth, they are actually of the order of 10^{-8} or smaller for the forced wobbles and about 10^{-6} for the free CW.) So we may approximate the Euler equations to the first order in these small quantities:

$$\begin{aligned} A\Omega_0 \frac{dm_x}{dt} + \Omega_0^2(C-B)m_y &= \Gamma_x \\ A\Omega_0 \frac{dm_y}{dt} + \Omega_0^2(A-C)m_x &= \Gamma_y \\ C\Omega_0 \frac{dm_z}{dt} &= \Gamma_z \end{aligned} \quad [12]$$

3.10.3.2.2 Axially symmetrical ellipsoidal case: Wobble motion

We specialize now to the Earth considered as an axially symmetrical, ellipsoidal, rigid body with no core.

The $(B-A)$ term in the third of the Euler equations [12] then drops out, and Γ_z vanishes too, because of axial symmetry. It follows that H_z remains constant, equal to $C\Omega_0$, that is, $m_z = 0$. Furthermore, with $(C-B)$ now equal to $(C-A)$, the addition of i times the second of the aforementioned equations to the first one yields a single equation involving the complex quantities:

$$\tilde{m} = m_x + im_y, \quad \tilde{\Gamma} = \Gamma_x + i\Gamma_y \quad [13]$$

The combined (complex) equation is

$$\frac{d\tilde{m}}{dt} - ie\Omega_0\tilde{m} = \frac{\tilde{\Gamma}}{A\Omega_0} \quad [14]$$

where e is the dynamical ellipticity in the sense in which this term is commonly used in the geophysics literature:

$$e = (C-A)/A \quad [15]$$

Solution of the equation in the preceding text is a trivial matter:

$$\tilde{m}(t) = e^{ie\Omega_0 t}\left[\tilde{m}(0) + \int \frac{\tilde{\Gamma}(t)}{A\Omega_0} e^{-ie\Omega_0 t}\, dt\right] \quad [16]$$

Free Wobble: The first term, $\tilde{m}(0)e^{ie\Omega_0 t}$, does not involve the torque. It is a free motion of the angular velocity vector around the figure axis with frequency $e\Omega_0$ or e cpsd, as may be seen readily by decomposing $\tilde{m}$ into its real and imaginary parts: $m_x = \tilde{m}(0)\cos e\Omega_0 t$, $m_y = \tilde{m}(0)\sin e\Omega_0 t$, assuming $\tilde{m}(0)$ to be real. It represents, in fact, the Euler free wobble, with e cpsd as its eigenfrequency. In the case of more realistic nonrigid Earth models, the free mode becomes the CW, with a frequency differing from $e\Omega_0$.

Forced Wobble: The second term in eqn [16] is the forced wobble, which can be evaluated once the time dependence of $\tilde{\Gamma}$ is known.

Let us consider a single spectral component of the torque, with frequency $\sigma\Omega_0$, that is, σ cpsd:

$$\tilde{\Gamma}(t) = \tilde{\Gamma}(\sigma)e^{i\sigma\Omega_0 t} \quad [17]$$

Then, eqn [14] simplifies to

$$i(\sigma - e)\tilde{m}(\sigma) = \tilde{\Gamma}(\sigma)/(A\Omega_0^2) \quad [18]$$

so that

$$\tilde{m}(\sigma) = -i\frac{\tilde{\Gamma}(\sigma)}{A\Omega_0^2(\sigma - e)} \quad [19]$$

The factor $1/(\sigma - e)$ represents the *resonance* in the forced wobble that is associated with the Euler free wobble having the eigenfrequency e cpsd. In later sections where we consider both rigid and nonrigid Earth models, we identify the rigid case by adding a subscript Rig (standing for 'Rigid'). Thus, $\tilde{m}$, m_x, etc., of this section would be denoted by $\tilde{m}_{\mathrm{Rig}}$, m_{Rig_x}, etc.

3.10.4 The Tidal Potential and Torque

Before going on to the treatment of more realistic Earth models, let us examine the gravitational effects of a celestial body B of mass M_B on the Earth.

3.10.4.1 Potential, Acceleration, and Torque

If the position vector of B relative to the geocenter is $\mathbf{r}_B$, then its gravitational potential at $\mathbf{r}$ in the Earth is

$$W_B(\mathbf{r}) = -\frac{GM_B}{|\mathbf{r}_B - \mathbf{r}|} = \frac{GM_B}{r_B}\left(1 - \frac{2\mathbf{r}\cdot\mathbf{r}_B}{r_B^2} + \frac{r^2}{r_B^2}\right)^{-1/2} \quad [20]$$

Since $r = |\mathbf{r}| \le a_e \approx 6378$ km, and the nearest solar system body is the Moon at a geocentric distance of about 400 000 km, $r/r_B < 0.016$, and therefore, an expansion in powers of (r/r_B) is useful:

$$W_B(\mathbf{r}) = -\frac{GM_B}{r_B}\left(1 + \frac{\mathbf{r}\cdot\mathbf{r}_B}{r_B^2} - \frac{r^2}{2r_B^2} + \frac{3(\mathbf{r}\cdot\mathbf{r}_B)^2}{2r_B^4} + \cdots\right) \quad [21]$$

The acceleration of a mass element at $\mathbf{r}$ is then

$$-\Delta W_B(\mathbf{r}) = -\frac{GM_B}{r_B}\left(\frac{\mathbf{r}_B}{2r_B^2} + \frac{3(\mathbf{r}\cdot\mathbf{r}_B)\mathbf{r}_B - r_B^2\mathbf{r}}{r_B^4} + \cdots\right) \quad [22]$$

The first term represents a uniform acceleration toward the external body. In case the body is the Sun or the Moon, it is this centripetal acceleration which keeps the Earth and the other body going in an orbit around each other. The second term in the acceleration is position-dependent and causes different mass elements in the Earth to move relative to one another to the extent allowed by the elastic forces counteracting it. The result is a deformation of the solid Earth, which is manifested as the *solid Earth tide*, and also a redistribution of water mass in the oceans, the *ocean tide*. The nonuniform acceleration is referred to as the tidal acceleration. The term in the potential W_B that gave rise to this tidal acceleration is designated as the *tidal potential* of degree 2. Succeeding terms in the expansion of W_B, which are not explicitly shown, constitute tidal potentials of degrees > 2. The torque on the whole Earth, which is responsible for the forced variations in Earth rotation, is

$$\boldsymbol{\Gamma} = \int -\rho(\mathbf{r})\mathbf{r} \times \Delta W_B(\mathbf{r})\, d^3r = \frac{GM_B}{r_B}\int \rho(\mathbf{r})\left(\frac{\mathbf{r}\times\mathbf{r}_B}{r_B^2} + \frac{3(\mathbf{r}\cdot\mathbf{r}_B)(\mathbf{r}\times\mathbf{r}_B)}{r_B^4} + \cdots\right) d^3r \quad [23]$$

where the integration is over the volume of the Earth. The first term in eqn [23] involves $\int \rho(\mathbf{r})\mathbf{r}\, d^3r$, which is simply the Earth's mass times the position vector of the center of mass; this vector is null because the center of mass is the origin (geocenter) itself. The torque is therefore given by the integral of the second term in the previous expression, apart from the much smaller higher-degree terms represented by the set of dots at the end.

3.10.4.2 Torque Components in Terrestrial Frame

We evaluate the components of the torque integral in a TRF by expressing $\mathbf{r}$ and $\mathbf{r}_B$ in terms of its components with respect to that frame. We choose the principal axis frame, which is defined as one in which the inertia tensor is diagonal. The x-axis of this frame is the principal axis of the least moment of inertia A; the longitude α of this axis in the ITRF is about $-14°.93$.

Since the elements C_{ij} of the tensor are defined by

$$C_{ij} = \int \rho(\mathbf{r})\left(r^2\delta_{ij} - x_i x_j\right) d^3\mathbf{r} \quad [24]$$

where δ_{ij} is the Kronecker delta function, we have, in the principal axis reference frame,

$$\begin{pmatrix} C_{11} \\ C_{22} \\ C_{33} \end{pmatrix} = \int \rho(\mathbf{r}) \begin{pmatrix} (y^2+z^2) \\ (z^2+x^2) \\ (x^2+y^2) \end{pmatrix} d^3 = \begin{pmatrix} A \\ B \\ C \end{pmatrix} \quad [25]$$

$$\int \rho(\mathbf{r}) xy\, d^3r = \int \rho(\mathbf{r}) yz\, d^3 = \int \rho(\mathbf{r}) zx\, d^3r = 0 \quad [26]$$

If the time-dependent variations in the density function $\rho(\mathbf{r})$ and hence in the inertia tensor due to tidal deformations and other causes were also taken into account, the result of the integrations in eqn [25] would get modified to $A + c_{11}$, $B + c_{23}$, $C + c_{33}$ and those in eqn [26] to $-c_{12}$, $-c_{23}$, $-c_{31}$, respectively, where the c_{ij} are the contributions from the deformations. They would lead to second-order contributions to the torque, which are very small; they will be ignored in the following. But they are not entirely negligible at the levels of accuracy required at present, and we shall refer to them again in Sections 3.10.9.1 and 3.10.9.6 on numerical modeling.

Now, we can express the factor $(\mathbf{r}\cdot\mathbf{r}_B)(\mathbf{r}\times\mathbf{r}_B)$ in the torque integral eqn [23] in terms of the components of $\mathbf{r}$ and $\mathbf{r}_B$ and then evaluate the integral using eqns [25] and [26]. Thus, we obtain

$$\begin{pmatrix} \Gamma_x \\ \Gamma_y \\ \Gamma_z \end{pmatrix} = \frac{3GM_B}{r_B^5}\begin{pmatrix} (C-B)y_B z_B \\ (A-C)x_B z_B \\ (B-A)x_B y_B \end{pmatrix} \quad [27]$$

Replacing the coordinates of the celestial body in terms of the polar coordinates $(r_B, \theta_B, \lambda_B)$, where θ_B is its colatitude and λ_B is its longitude in the terrestrial frame that we are using, we rewrite the equatorial components of the torque as

$$\Gamma_x = \frac{3GM_B}{r_B^3}\bar{A}(e - e') \times \cos\theta_B \sin\theta_B \sin\lambda_B$$
$$\Gamma_y = \frac{3GM_B}{r_B^3}\bar{A}(e + e') \times \cos\theta_B \sin\theta_B \cos\lambda_B \quad [28]$$

wherein $\bar{A} = (A + B)/2$, and

$$e = (C - \bar{A})/\bar{A}, \quad e' = (B - A)/(2\bar{A}) \quad [29]$$

The parameter e', proportional to $(B - A)$, is a manifestation of the triaxiality of the Earth, and e as defined here is the dynamical ellipticity of the triaxial Earth; the latter dominates the torque because e' is only about $e/300$.

The complex combination $\tilde{\Gamma} = \Gamma_x + i\Gamma_y$ is

$$\tilde{\Gamma} = -i\frac{3GM_B}{r_B^3}\bar{A}\cos\theta_B \sin\theta_B\left(e e^{i\lambda_B} + e' e^{-i\lambda_B}\right) \quad [30]$$

Now, as seen from the rotating Earth, all celestial objects move westward in sky, that is, in a retrograde sense (opposite to the eastward rotation of the Earth itself, as a consequence of this rotation). Their angular velocities $d\lambda_B/dt$ are close to -1 cpsd. In

fact, the temporal variation of the torque due to the orbital motion of the Sun or the Moon has spectral components with frequencies on both sides of −1 cpsd even if the orbits are taken to be circular, as will be demonstrated in the next couple of sections. The fact that the orbits are elliptical and that even these are perturbed by the influence of third bodies (e.g., perturbation of the lunar orbit around the Earth by the attraction of the Sun and the planets on the Moon) leads to a rich spectrum of frequencies bunched very close to and centered at −1 cpsd. This band of frequencies is referred to as the *retrograde diurnal band*. An examination of eqn [28] shows now that the main part proportional to $(C-\bar{A})$ of the equatorial projection (Γ_x, Γ_y) of the torque vector rotates in the retrograde sense as a consequence of the negative rate of change of λ_B and that this motion has the same retrograde diurnal spectrum; the triaxiality part of the vector, on the other hand, rotates in the prograde sense and has a prograde diurnal spectrum.

Another point of interest is that the products $y_B z_B$ and $x_B z_B$ present in Γ_x and Γ_y in eqn [27] come from the terms

$$-\frac{3GM_B}{r_B^5}(x_B z_B xz + y_B z_B yz) = -\Omega_0^2(\phi_1 xz + \phi_2 yz) \qquad [31]$$

of the potential eqn [21]. The dimensionless quantities ϕ_1, ϕ_2 and their complex combination $\tilde{\phi} = \phi_1 + i\phi_2$ were introduced by Sasao et al. (1980) and have been used extensively in further developments based on their formalism. A potential that is a linear combination of xz and yz is called a *tesseral* potential of degree 2. Thus, the equatorial components of the gravitational torque due to the degree 2 terms in the potential are produced solely by the tesseral part of these terms, and the only Earth parameters on which this torque depends are the differences among the three principal moments of inertia, which, in turn, depend on the dynamical ellipticity e and the triaxiality (fractional difference between the two equatorial moments of inertia). Since higher-degree terms in the expansion of the full potential are of much smaller magnitude as noted earlier, the degree 2 tesseral potential, which has its spectrum in the retrograde diurnal band, is responsible for the dominant part of the wobbles and hence of the nutations too, as we shall see. The residual small part of the wobble and nutation arise from the action of potentials of degree >2 on moments of the Earth's matter distribution that are of degree >2 (the moments of inertia being of degree 2).

3.10.4.3 Expansion of the Potential: Tidal Spectrum

The terms up to degree 2 in (r/r_B) in the expansion of the potential $W_B(\mathbf{r})$ of an external body B were shown in eqn [21]. A further expansion of each of the tidal terms (of degree $n \geq 2$) can be made into parts of different orders m from 0 to n, which are distinguished by the nature of their dependence on the direction of the position $\mathbf{r}_B$ of the body B and that of the Earth point $\mathbf{r}$ relative to the terrestrial reference frame (ITRF). This dependence is through the surface spherical harmonic functions Y_n^m of the colatitude θ and longitude λ of $\mathbf{r}$ in the ITRF, as well as the Y_n^m of θ_B and λ_B of $\mathbf{r}_B$ in the same frame. (Note that λ and λ_B in this frame are equal to α plus the corresponding quantities relative to the principal axis frame of the last section.)

The spherical harmonics are defined by

$$Y_n^m(\theta,\lambda) = N_{nm} P_n^m(\cos\theta) e^{im\lambda} \qquad [32]$$

and

$$Y_n^{-m} = (-1)^m Y_n^{m^*} \qquad [33]$$

with * standing for complex conjugation and with

$$N_{nm} = (-1)^m \left(\frac{2n+1}{4\pi}\right)^{1/2} \left(\frac{(n-m)!}{(n+m)!}\right)^{1/2} \qquad [34]$$

for $n = 0, 1, 2, \ldots, m = 0, 1, \ldots, n$. P_n^m are the associated Legendre functions; they are derived from the Legendre polynomials P_n, which are defined through the expansion

$$(1 - 2cw + w^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(c) w^n \qquad [35]$$

$$P_n^m(c) = (1-c^2)^{m/2} \frac{d^m P_n(c)}{dc^m} \qquad [36]$$

$P_n(c)$ is a polynomial of degree n in c and has unit value at $c = 1$. The spherical harmonics Y_n^m for all n and $m = -n, -n+1, \ldots, n-1, n$, form an orthonormal set:

$$\int Y_q^{p^*}(\theta,\lambda) Y_n^m(\theta,\lambda) \sin\theta \, d\theta \, d\lambda = \delta_{qn}\delta_{pm} \qquad [37]$$

The integration here is over the unit sphere and δ_{qn} is the Kronecker delta function: δ_{qn} is 1 if $q = n$ and is 0 otherwise. The expansion of the potential eqn [20] in terms of the Y_m^n is, after dropping terms of degrees $n = 1, 2$, which do not produce any tidal deformations or torques,

$$W_B(\mathbf{r}) = -\sum_{n=2}^{\infty} r^n \sum_{m=0}^{n} [F_{nm}^*(r_B, , \theta_B, , \lambda_B) Y_n^m(\theta,\lambda) + F_{nm}(r_B, \theta_B, \lambda_B) Y_n^{m^*}(\theta,\lambda)] \qquad [38]$$

with

$$F_{nm}(r_B, \theta_B, \lambda_B) = \frac{GM_B}{r_B^{n+1}} \frac{(2-\delta_{m_0})2\pi}{2n+1} \times Y_n^m(\theta_B, \lambda_B) \qquad [39]$$

and the complex conjugate of it provides the expression for $F_{nm}^*(r_B, \theta_B, \lambda_B)$. It may be observed that the part of W_B belonging to any specific degree n and order m contains the phase factors $e^{im(\lambda-\lambda_B)}$ and $e^{im(\lambda-\lambda_B)}$. Each of the phases is 0 on the meridian $\lambda = \lambda_B$, and as the celestial bodies move westward in the sky (i.e., in the retrograde sense) at a rate close to once per day, as seen from the terrestrial frame, the previously mentioned meridian moves with it; other constant phase meridians move in tandem. So, for any $m \neq 0$, this term constitutes a retrograde tidal wave sweeping over the volume of the Earth, with m cycles over the globe (because of the factor m multiplying λ).

Representation of the time dependence of the potential through its spectral expansion is of great value. The most widely used form is that of the expansion by Cartwright and Taylor (1971), though a number of tidal tables of much higher accuracy are now available (Hartmann and Soffel, 1994; Hartmann and Wenzel, 1995a,b; Hartmann et al., 1999; Roosbeek, 1999; Roosbeek and Dehant, 1998):

$$F^{*}_{nm}(r_B,\theta_B,\lambda_B) = (g_e/a_e^n)\sum_{\omega} H^{nm}_{\omega}\zeta_{nm}e^{i\Theta_\omega(t)} \qquad [40]$$

wherein the amplitude of the spectral component is given in units of length as the equivalent 'height' H^{nm}_{ω}. The factor ζ_{nm} is defined to be unity when $n-m$ is even and $-i$ when $n-m$ is odd. With these values for ζ_{nm}, the dependence of the tidal component on the longitude λ and time appears through the factor $\cos(\lambda+\Theta_\omega(t))$ for $(n-m)$ even and $\sin(\lambda+\Theta_\omega(t))$ for $(n-m)$ odd, when the real part is taken in eqn [38].

$\Theta_\omega(t)$ is the argument of the tidal component; the frequency ω is nonnegative and is equal to $d\Theta_\omega/dt$:

$$\Theta_\omega(t) = \omega t + \mathcal{X}_\omega \qquad [41]$$

where χ_ω is a constant. $\Theta_\omega(t)$ does have a dependence on n and m, which is not indicated explicitly. Now, the mean motion of B in the sky (i.e., the mean value of $d\lambda_B/dt$) relative to the terrestrial frame is close to -1 cpsd, and $e^{i\Theta_\omega(t)}$ for given n and m originates from $e^{-im\lambda_B}$ (modulated by other factors depending on θ_B and r_B), as is evident from the expression for F^{*}_{nm}. Hence, the frequency spectrum of $e^{i\Theta\omega(t)}$ must lie in a band around m cpsd. Since the angle of a spectral component of the wave of order m is now $(\lambda+\Theta_\omega(t))$, it constitutes a retrograde wave with frequency $-\omega$ lying within the retrograde band around $-m$ cpsd.

The roles of the various solar system bodies in the argument $\Theta_\omega(t)$ are displayed through its expansion in terms of Doodson fundamental arguments of the tides:

$$\Theta_\omega = n_1(\tau-\lambda) + n_2 s + n_3 h + n_4 p + n_5 N' + n_6 p_s + \cdots \qquad [42]$$

where $n_1 = m$ and the set of dots stand for terms that are multiples of the mean tropic longitudes of the planets from Mercury to Saturn. Of the fundamental arguments that are explicitly shown, s, h, p, $-N$, p_s are the mean tropic longitudes of the Moon, Sun, Moon's perigee, Moon's node, and the Sun's perigee, respectively. The periods relating to these arguments range from 27.32 days for s to about 209 years for p_s. The Doodson argument τ is $180^\circ + \text{GMST} - s + \lambda$ when expressed in degrees. GMST is the Greenwich Mean Sidereal Time, which is the time measured by the angle traversed along the mean equator of date by the Greenwich meridian at the mean sidereal rotation rate Ω_0:

$$\text{GMST} = \text{GMST}_0 + \Omega_0 t + \cdots \qquad [43]$$

$$\text{GMST}_0 = 280^\circ.4606\cdots \qquad [44]$$

where t is measured from J2000. The period associated with $\tau-\lambda+s$ is that of the mean sidereal rotation, namely, 1 sidereal day. It is the term $n_1\tau = m\tau$ in Θ_ω that causes the frequencies in potentials of order m to lie within a band centered at m cpsd.

3.10.4.4 Degree 2 Potential and Torque: Spectra

The tidal torque on an ellipsoidal Earth arises from the degree 2 part of the potential W_B, as has been noted earlier. The degree 2 potential, in turn, is made up of zonal, tesseral, and sectorial parts, which are of orders $m=0$, 1, and 2, respectively. The associated Legendre functions appearing in the three orders are

$$\begin{aligned} P_2(\cos\theta) &= \frac{3\cos^2\theta - 1}{2} \\ P^1_2(\cos\theta) &= 3\sin\theta\cos\theta \\ P^2_2(\cos\theta) &= 3\sin^2\theta \end{aligned} \qquad [45]$$

These are factors in the spherical harmonics $Y^m_2(\theta,\lambda)$ and appear with the coefficients $F^{*}_{2m}(\theta_B,\lambda_B)$ in the tidal potential. Their time dependence enters through $e^{i\Theta_\omega(t)}$ with $n_1 = m = 0$, 1, 2, respectively, as we see from eqns [40] and [42]. Therefore, the spectra of the zonal, tesseral, and sectorial potentials lie in the low-frequency, prograde diurnal, and prograde semidiurnal bands, respectively, as will be clear from the discussion at the end of the last section.

3.10.4.4.1 Equatorial components of torque

We can now carry over $\tilde{\Gamma}$ of eqn [30], which was with respect to the principal axis frame, to the ITRF. In doing so, we have to make the replacement $\lambda_B \to (\lambda_B - \alpha)$, as the x-axis of the principal axis frame is at longitude α in the ITRF as already noted. Secondly, the transformation of the vector components between the two frames makes $\tilde{\Gamma}$ in the ITRF $e^{i\alpha}$ times $\tilde{\Gamma}$ of the other frame though Γ_z remains unchanged. In consequence of these two changes, eqn [30] goes over into

$$\begin{aligned} \tilde{\Gamma} &= -i\frac{GM_B}{r_B^3}P^1_2(\cos\theta_B)\bar{A}\left(e e^{i\lambda_B} + e' e^{2i\alpha}e^{-i\lambda_B}\right) \\ &= i\frac{GM_B}{r_B^3}\left(\frac{24\pi}{5}\right)^{1/2}\bar{A}\times\left(eY^1_2(\theta_B,\lambda_B) + e' e^{2i\alpha}Y^{1*}_2(\theta_B,\lambda_B)\right) \end{aligned} \qquad [46]$$

where the definitions [32] and [34] have been used in the second step. One sees immediately by comparison with eqn [40] that the ellipticity term in the last expression is F^{*}_{21} apart from constant factors and that the triaxiality term is proportional to F_{21}. Then, eqn [40] taken together with the ensuing discussion shows that the spectra of these two terms lie in retrograde diurnal and prograde diurnal bands, respectively.

We have observed earlier that it is the tesseral part which produces the equatorial torque on an ellipsoidal Earth; this torque drives the wobbles. The form employed in eqn [31] for the tesseral potential may be related readily to F_{21}. We note that

$$-\Omega_0^2(\phi_1 xz + \phi_2 yz) = -\frac{\Omega_0^2 r^2}{3}\text{Re}\left(\tilde{\phi}^*(t)P^1_2(\cos\theta)e^{i\lambda}\right) \qquad [47]$$

where $\tilde{\phi} = \phi_1 + i\phi_2$. Comparison with the $n=2$, $m=1$ term in eqn [38] after writing $P_{21}e^{i\lambda}$ as $-(24\pi/5)^{1/2}Y_{21}$ enables us to identify the dimensionless quantity $\tilde{\phi}^{*}$ as

$$\tilde{\phi}^*(t) = -\left(\frac{15}{8\pi}\right)^{1/2}\frac{F^{*}_{21}}{\Omega_0^2} \qquad [48]$$

We turn our attention now to the equatorial torque $\tilde{\Gamma}$. One can express it in terms of F_{21} and F^{*}_{21} by using the definition of Y^1_2 from eqn [32]. One finds, after some elementary algebra, that

$$\tilde{\Gamma} = \left(\frac{15}{8\pi}\right)^{1/2} i\bar{A}\left(eF_{21} + e' e^{2i\alpha}F^{*}_{21}\right) \qquad [49]$$

An alternative expression for $\tilde{\Gamma}$ in terms of $\tilde{\phi}$ may be obtained by using eqn [48]:

$$\tilde{\Gamma} = -iA\Omega_0^2\left(e\tilde{\phi} + e'e^{2i\alpha}\tilde{\phi}^*\right) \quad [50]$$

This form, especially for the ellipticity part, is widely used in works following the approach of Sasao et al. (1980).

We write down now the spectral expansion of $\tilde{\phi}$ by making use of the relation [48] and the expansion of F_{21}^* from eqn [40]:

$$\tilde{\phi} = -i\left(\frac{15}{8\pi}\right)^{1/2}\frac{g_e}{a_e^2\Omega_0^2}\sum_\omega H_\omega^{21}e^{-i\Theta_\omega(t)} \quad [51]$$

The expansion of the torque follows immediately. An important point to note is that the ellipticity part of the torque involves $\tilde{\phi}$ with the time dependence $e^{-\Theta_\omega(t)} = e^{-i(\omega t+\chi_\omega)}$ for the spectral terms, so that the spectral frequencies are negative (since ω is positive); when combined with the fact that $m=1$ for the tesseral potential, it follows that the spectrum is in the retrograde diurnal band.

In the work of Sasao et al. (1980), and in many other related works, it is customary to use the notation $\sigma\Omega_0$ for the (negative) spectral frequencies involved and $\tilde{\phi}(\sigma)$ for the amplitude of the spectral component. With this notation, a typical component of $\tilde{\phi}$ is given by

$$\tilde{\phi}_\sigma(t) = -i\tilde{\phi}(\sigma)e^{i(\sigma\Omega_0 t+\chi_\sigma)} \quad [52]$$

with

$$\sigma = -\omega/\Omega_0, \quad \chi_\sigma = -\chi_\omega \quad [53]$$

with the amplitude $\tilde{\phi}(\sigma)$ related to the height H_ω^{21} of Cartwright and Taylor (1971) by

$$\tilde{\phi}(\sigma) = \left(\frac{15}{8\pi}\right)^{1/2}\frac{g_e}{a_e^2\Omega_0^2}H_\omega^{21} \quad [54]$$

Finally, we obtain an explicit expression for the wobble amplitude $\tilde{m}_R(\sigma)$ of a rigid axially symmetrical Earth by introducing the ellipticity part of the torque eqn [50] into the solution [19] of the torque equation:

$$\tilde{m}_R(\sigma) = -\frac{e\tilde{\phi}(\sigma)}{\sigma - e} \quad [55]$$

Tables of the components of the tidal potential list the values of H_ω for the various components identified by the respective sets of values of the multipliers $n_1, n_2, \ldots, n_6$ of the Doodson arguments. The amplitude $\tilde{\phi}(\sigma)$ of any tide of interest may be computed using eqn [54] after picking out the H_ω of that tide from a tide table. The rigid Earth wobble amplitude may then be computed from eqn [55], given the value of e.

3.10.4.4.2 Axial component of the torque

The z component of the torque, given in eqn [27], may be rewritten with the replacement $\lambda_B \to (\lambda_B - \alpha)$, as

$$\begin{aligned}\Gamma_z &= -i\frac{3GM_B}{4r_B^3}(B-A)\sin^2\theta_B\left(e^{2i(\lambda_B-\alpha)} - e^{-2i(\lambda_B-\alpha)}\right)\\ &= -i\frac{GM_B}{r_B^3}(B-A)\left(\frac{6\pi}{5}\right)^{1/2}\left[e^{-2i\alpha}Y_2^2 - e^{2i\alpha}Y_2^{2^*}\right]\end{aligned} \quad [56]$$

This translates into

$$\Gamma_z = -i\left(\frac{15}{8\pi}\right)^{1/2}\bar{A}e'\left(e^{-2i\alpha}F_{22} - e^{2i\alpha}F_{22}^*\right) \quad [57]$$

The spectral expansion of this torque may be obtained as a special case of eqn [40]. Since $m=\pm 2$ in the present case, the two terms within parentheses have their spectra in the retrograde and prograde semidiurnal bands, respectively. The amplitudes are very small because e'/e is only about 1/300 as already noted.

3.10.4.5 Terms of General Degree and Order in the Torque

The torque generated by any higher-degree term in the potential may be obtained by taking $(-\mathbf{r}\times\nabla)$ of the (nm) part of eqn [38], then multiplying by $\rho(\mathbf{r})$, and integrating over the volume of the Earth. On using the fact that

$$\left[(\mathbf{r}\times\nabla)_x + i(\mathbf{r}\times\nabla)_y\right]Y_n^m = i[(n-m)(n+m+1)]^{1/2}Y_n^{m+1} \quad [58]$$

for any m, positive or negative, and that $Y_n^{m*} = (-1)^m Y_n^{-m}$, it turns out that the equatorial part of the torque is given by

$$\begin{aligned}\tilde{\Gamma}^{nm} = iM_E a_e^n\frac{2n+1}{16\pi N_{nm}}\Big\{&(C_{n,m-1} - iS_{n,m-1})\frac{2F_{n,m}}{(2-\delta m,1)}\\ &-(n+m)(n-m+1)(C_{n,m+1} + iS_{n,m+1})F_{n,m}^*\Big\}\end{aligned} \quad [59]$$

where C_{nm} and S_{nm} are geopotential coefficients defined by

$$\begin{aligned}&M_E a_e^n(C_{n,m} + iS_{nm})\\ &\quad = (2-\delta_{m,0})\frac{(n-m)!}{(n+m)!}\int\rho(\mathbf{r})r^n P_n^m(\cos\theta)e^{im\lambda}\,d^3r\end{aligned} \quad [60]$$

Here, M_E is the mass and a_e the equatorial radius of the Earth. It may be seen from the spectral representation [40] of F_{nm}^* and its complex conjugate that the spectra of the two terms of eqn [59] lie, in general, in the two bands centered at $-m$ cpsd and m cpsd, respectively. (In the special case $m=0$, it turns out that both parts of the torque involve $(C_{n1}+iS_{n1})$ and that the spectra of the two parts together make up the band of width 1 cpsd centered at $m=0$.)

The coefficients C_{n0} of order 0, also denoted by $-J_n$, represent axially symmetrical structures. They appear only through the first term of eqn [59] with $m=1$. The torque resulting from the existence of C_{n0} may be seen to reduce to

$$i\frac{GM_B M_E}{r_B}\left(\frac{a_e}{r_B}\right)^n C_{n0}P_{n1}(\cos\theta_B)e^{i\lambda_B} \quad [61]$$

(Note that all S_{n0} vanish.)

The axially symmetrical ellipsoidal Earth is a special case with $n=2$, for which

$$C_{20} = -J_2 = -\left(\bar{A}/M_E a_e^2\right)e \quad [62]$$

It is of some interest to note that the triaxiality enters through

$$\begin{aligned}C_{22} &= \left(\bar{A}/2M_E a_e^2\right)e'\cos 2\alpha\\ S_{22} &= \left(\bar{A}/2M_E a_e^2\right)e'\sin 2\alpha\end{aligned} \quad [63]$$

In this context, α stands for the longitude of the direction, in the ITRF, of the equatorial axis of the least moment of inertia A. (This angle was introduced in Section 3.10.4.2. It is to be kept in mind that α is used elsewhere to denote the right ascension.)

The z component of the general torque may also be obtained by making use of the identity $(\mathbf{r}\times\nabla)_z Y_n^m = imY_n^m$. One finds that

$$\Gamma_z^{nm} = -iM_E a_e^n \frac{2n+1}{4\pi}\frac{m}{N_{nm}(2-\delta_{m0})} \times\left\{(C_{nm}-iS_{nm})F_{nm} - (C_{n,m}+iS_{n,m})F_{n,m}^*\right\} \quad [64]$$

3.10.5 Torque in Celestial Frame: Nutation–Precession in a Simple Model

3.10.5.1 Torque Components in Celestial Frame

We considered in the last section the torque and rotation variations referred to the TRF. One might ask: What if we employed the components (X, Y, Z) of $\mathbf{r}$ and (X_B, Y_B, Z_B) of $\mathbf{r}_B$ in a space-fixed equatorial reference frame (say, the GCRF) and wished to compute $(\Gamma_X, \Gamma_Y, \Gamma_Z)$ in that frame on the same lines as in the foregoing section? The problem that one encounters in attempting such a course is that neither the Earth's z-axis (axis of maximum moment of inertia) nor the Earth's rotation vector stays aligned with the Z-axis of the celestial frame, as a result of precession–nutation and wobble. However, we may ignore the small time-dependent offsets among these axes for the limited purpose of a lowest-order calculation of the torque, and then, if we also ignore the small deviation of the Earth from axial symmetry, the axial symmetry persists about the Z-axis too. Considering these approximations, one may evaluate the integral [23] in the space-fixed frame by expressing the integrand in terms of coordinates (X, Y, Z) and (X_B, Y_B, Z_B) and noting that the results of eqns [25] and [26] remain valid with these replacements. The components of $\boldsymbol{\Gamma}$ in the space-fixed frame are then found to be

$$\Gamma_X = \frac{3GM_B}{r_B^5}(C-A)Y_B Z_B$$
$$\Gamma_Y = -\frac{3GM_B}{r_B^5}(C-A)Z_B X_B \quad [65]$$

with $\Gamma_Z = 0$.

The values of r_B, X_B, Y_B, Z_B as functions of time may be taken, for any of the celestial bodies, from the appropriate ephemerides such as VSOP (Variations Séculaires des Orbites Planétaires) for the Sun and the planets (Bretagnon and Francou, 1988) and ELP (Éphéméride Lunaire Parisienne) for the Moon (Chapront-Touzé and Chapront, 1988). The time dependence of the torque components is then known. The nutation of the angular momentum axis may then be obtained by direct integration in view of eqns [6] and [7]. An interesting fact that is of importance is that though the orbital motions of both the Sun and the Moon, as seen from a nonrotating geocentric frame, are in the prograde sense relative to Earth's rotation, the torque exerted by each of the bodies contains both prograde and retrograde components relative to such a frame. A treatment based on a simplified model of the orbital motion of the Sun will bring out clearly why this happens, besides providing a concrete example of the calculation of nutation and precession.

3.10.5.2 Simple Model: Sun in a Circular Orbit

We consider an axially symmetrical ellipsoidal Earth subjected to the tidal potential of the Sun, which is supposed, for illustrative purposes, to be in uniform motion with angular velocity p along a circle of radius c centered at the geocenter and lying in the plane of the ecliptic. We take the space-fixed reference frame to be equatorial celestial frame (GCRF), with its X-axis pointing in the direction of the vernal equinox. (This is the direction of increasing ecliptic longitude, that is, of positive $\Delta\psi$. The motion in obliquity of the figure axis is in the direction of the Y-axis.) The angle between the principal plane and the ecliptic is of course the obliquity ε, which we shall treat as a constant. Then, the coordinates of the Sun in our reference system are

$$X_B = c\cos pt, \quad Y_B = c\cos\varepsilon\sin pt, \quad Z_B = c\sin\varepsilon\sin pt \quad (66)$$

Note that (X_B, Y_B) describes the projection of the Sun's circular orbit from the ecliptic plane on to the equatorial plane; the projected orbit is, not surprisingly, an ellipse.

3.10.5.3 The Torque in the Simple Model

The time dependence of the torque eqn [65] can now be displayed in explicit form:

$$\Gamma_X = F\cos\epsilon(1-\cos 2pt)$$
$$\Gamma_Y = -F\sin 2pt, \quad \Gamma_Z = 0 \quad [67]$$

where

$$F = \left(3GM_B/2c^3\right)Ae\sin\epsilon \quad [68]$$

We know from Kepler's laws that $GM_B = n_B^2 c^3$, where n_B is the 'mean motion,' that is, the mean orbital angular velocity of the body, which is what we denoted by p earlier. (In the case of an elliptical orbit, c has to be the semimajor axis of the orbit in Kepler's relation.)

Thus, the overall magnitude of the solar torque on the Earth is determined by p and the obliquity ε, besides the Earth parameter $(C-A)$. One may get an idea of the strength of the torque from the value of F. For the rotation variations induced by the torque, the nondimensional quantity $F/(\Omega_0^2 A)$ is more relevant:

$$\frac{F}{\Omega_0^2 A} = \frac{3}{2}\frac{p^2}{\Omega_0^2}e\sin\epsilon \quad [69]$$

We know that $p/\Omega_0 = 1/366.2422$, since the period of the relative orbital motion of the Sun and the Earth is 1 year of 365.2422 solar days or 366.2422 sidereal days. We also have, from the IERS Conventions 2000 (McCarthy, 2003), $\sin\varepsilon = 0.397777$ and $\cos\varepsilon = 0.917482$. Therefore, $F/(\Omega_0^2 A) = 4.449\times10^{-6}e$; it is a dimensionless measure of the amplitude of the torque. On using the value $e = 0.0032845$ from a recent estimate, one finds it to be 1.461×10^{-8}.

Returning to eqns [67] and [69], we take note of three important features that are evident: the presence of a constant (time independent) term in the X component of the torque, the semiannual frequency of the periodic part of both the equatorial components of the torque, and the fact that the periodic part of the variation of (Γ_x, Γ_y) is an elliptical motion in the equatorial plane.

3.10.5.4 Nutation and Precession in the Model

With the components of $\boldsymbol{\Gamma}$ given in the celestial frame by eqn [67], explicit computation of the motion of the angular

momentum axis in space may be done trivially using the eqn [5] according to which $X_H = \varepsilon\Gamma_X\, dt$ and $Y_H = \varepsilon\Gamma_Y\, dt$. We find thus that

$$X_H = \frac{3p^2}{2\Omega_0} H_d \sin\varepsilon\cos\varepsilon[t - (1/2p)\sin 2pt]$$
$$Y_H = \frac{3p^2}{2\Omega_0} H_d \sin\varepsilon[(1/2p)\cos 2pt] \qquad [70]$$

where $H_d = (C-A)/C = e/(1+e)$. The periodic parts of X_H and Y_H evidently describe an *elliptical* nutation.

Now, we resolve the two-vector with components X_H, Y_H into two parts as follows:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = D\left[(1+\cos\varepsilon)\begin{pmatrix} -\sin 2pt \\ \cos 2pt \end{pmatrix} + (1-\cos\varepsilon)\begin{pmatrix} \sin 2pt \\ \cos 2pt \end{pmatrix}\right] \qquad [71]$$

where $D = (3p/8\Omega_0)\, H_d \sin\varepsilon$. The first vector on the right-hand side rotates in the prograde sense (the same as the rotation of the Earth, from the positive direction of the X-axis toward that of the Y-axis), and the second one's rotation is retrograde. Thus, the aforementioned equation constitutes a resolution of the elliptical motion of the angular momentum axis into prograde and retrograde circular motions. The factor $(1-\cos\varepsilon)$ in the amplitude of the retrograde component makes it clear that its presence (despite the motion of the Sun being in the prograde sense in the ecliptic plane) is because of the nonzero obliquity of the ecliptic relative to the equator. It is another matter that D becomes zero and therefore the torque as a whole vanishes if $\varepsilon = 0$.

The complex combination of X_H and Y_H is

$$X_H + iY_H = \mathrm{i}(3p/8\Omega_0)H_d \sin\varepsilon \times \left[(1+\cos\varepsilon)e^{2ipt} + (1-\cos\varepsilon)e^{-2ipt}\right] \qquad [72]$$

The prograde part is characterized by a positive frequency $(2p)$ and the retrograde part by a negative frequency $(-2p)$. This is a very general property, as will be seen in the next section.

It is worth emphasizing here that the results [70] and [72] (and also eqn [73]) are for the angular momentum axis and not for the figure axis. These results are independent of whether the Earth is rigid or nonrigid and of whether or not it has a core, since there was no need to input any information about the structure and properties of the Earth (other than that of axial symmetry) in order to obtain the solution for the angular momentum vector. In contrast, the motions of the figure axis and the rotation axis are strongly influenced by the Earth's properties. As a consequence, studies on those motions are the ones that serve to provide glimpses into the properties of the Earth's interior.

We return now to eqn [66] and rewrite their periodic parts in terms of the classical nutation variables $\Delta\psi_H$ and $\Delta\varepsilon_H$ using eqn [1]:

$$\Delta\psi_H = -(3p/4\Omega_0)H_d \cos\epsilon \sin 2pt$$
$$\Delta\epsilon_H = (3p/4\Omega_0)H_d \sin\epsilon \cos 2pt \qquad [73]$$

The coefficients of $\sin 2pt$ and $\cos 2pt$ in these expressions are known as the *coefficients of nutation* in longitude and obliquity.

For this semiannual nutation of angular frequency $2p = 2n_s = 4\pi$ rad year^{-1}, one finds the coefficients in longitude and latitude to be -1269 and 550 mas, respectively (this corresponds to a quasi-circular motion at the Earth's surface of about 16 m as seen from space). All these numbers are very close to those obtained from accurate treatments of the problem. The largest of the nutation terms is of lunar origin. It has a period of about 18.6 years, which arises from the precession of the lunar orbit around the ecliptic. The coefficients in longitude and latitude of this nutation are about $-17\,000$ mas and 9200 mas, respectively. A representation of the nutational motion is provided in Figures 1, 2, 3, and 6.

As for the term in X_H, which is linear in t, it represents precession in longitude. The precession rate is

$$\dot{\psi}_{A,H} = \frac{1}{\sin\epsilon}\frac{dX_{H\,sec}}{dt} = (3p^2/2\Omega_0)H_d\cos\epsilon \qquad [74]$$

On using the values given earlier for the various parameters, one finds that the precession rate due to solar gravitational torque as given by eqn [74] is 7.7291×10^{-5} rad year^{-1} or, equivalently, 15 943 mas year^{-1}. Lunar attraction too gives rise to precessional motion, a little more than double that due to the Sun; the greater role of the Moon despite its mass being very much smaller than that of the Sun is a consequence of its being very much closer to the Earth. The total precession rate is about 50 000 mas year^{-1}.

3.10.6 Elliptical Motions: Prograde and Retrograde Circular Components

We have observed in the previous example that the periodic part of the motion of the vector (X_H, Y_H) describes an elliptical path. Periodic motions of the figure axis or the rotation axis in space are also elliptical. For our purposes, the resolution of the elliptical nutations of the axes and of torque vectors responsible for the nutations into pairs of counterrotating circular components is important, because the Earth's nutational responses to the two circular components are far from similar. In the following, we consider specifically the elliptical motions in the celestial frame; but it may be kept in mind that the resolution into circular components can be carried out equally well for motions, such as the wobble, in the terrestrial frame (and indeed, for any elliptical motion in general).

3.10.6.1 The Spectrum and the Fundamental Arguments of Nutation

The example of the last section is a gross oversimplification of the orbital motion of the Sun as seen from the GCRF. The actual orbit is elliptical and not circular, and even the elliptical orbit is perturbed by the Moon's pull on the Earth and other influences; consequently, the orbital motion has a spectrum of frequencies, which will be reflected in the spectrum of the solar torque on the Earth. The contributions from the orbital motions of the Moon and the Sun (perturbed by the planetary influences) make up all but a small part of the spectrum; the remainder consists of the relatively minor but nonignorable contributions from the planets. An argument $\Xi(t)$ of the sine and cosine terms in a typical spectral component of the torque bears therefore the imprint of more than one of these bodies and is expressed as a linear combination, with integer coefficients, of a number of *fundamental arguments*, each of which

relates to the motion of one or the other of the solar system bodies. The fundamental arguments relating to the lunar and solar motions are the Delaunay arguments. They are

- ℓ = mean anomaly of the Moon,
- ℓ' = mean anomaly of the Sun,
- $F = L - \Omega$,
- D = mean elongation of the Moon from the Sun,
- Ω = mean longitude of the ascending node of the Moon,

where L is the mean longitude of the Moon.

Additional fundamental arguments relating to the planets are $\lambda_1, \lambda_2, \ldots, \lambda_8$, which are the mean longitudes with respect to the fixed equinox and ecliptic of J2000, in a solar system barycentric reference frame, of Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune, respectively. Each of these planetary arguments is taken to be linear in time. One more argument that appears in planetary terms in nutation is the precession angle p_a, in which a t^2 term is also retained.

The argument of a general term is then

$$\Xi(t) = n_1 l + n_2 l' + n_3 F + n_4 D + n_5 \Omega + \sum_{i=1}^{9} n_i \lambda_i \qquad [75]$$

where $\lambda_9 = p_a$. The frequency of this spectral term is $d\Xi/dt$.

A listing of the fundamental arguments as functions of time may be found in the IERS Conventions 2000 (McCarthy, 2003). The periods corresponding to the frequencies of the lunar and solar arguments l, l', F, D, and Ω are 27.5545, 365.2596, 27.2122, 29.5306, and −6798.384 solar days, respectively. The last of these periods, which is about 18.6 years, is the period of the precession of the orbital plane of the Moon around the normal to the ecliptic. The period corresponding to L is 27.3216 solar days; it is the period of revolution of the Moon in its orbit. All these arguments are very nearly but not exactly linear in t, meaning that the periods are very slowly varying with time. The periods used to identify them are the values pertaining to J2000. The periods of the planets, which relate to the arguments λ_i ($i = 1, 2, \ldots, 8$), range from about 88 days for Mercury to about 60 000 days for Neptune, and the period relating to p_a is about 25 770 years. It is clear then that the frequencies of the fundamental arguments are much smaller than once per day; all the $\Xi(t)$ of interest to the nutations of an axially symmetrical ellipsoidal Earth have the same property. Thus, they are all in the low-frequency band, that is, between −0.5 and 0.5 cpsd.

Bretagnon et al. (1997) observed that ℓ' and Ω are so close to $2l - 2D - \lambda_3$ and $\lambda_3 + D - F$, respectively, that observations over periods of the order of thousands of years would be needed to distinguish them. Therefore, in developing their nutation series, they dropped ℓ and Ω from the fundamental set, replacing them by the aforementioned combinations plus small corrections that are linear in t; they also dropped the argument p_a, which has a very long period. Furthermore, the highly accurate nutation series that they have developed in terms of the truncated set of 11 fundamental arguments are not for the classical $\Delta\psi$ and $\Delta\varepsilon$, which are referred to the moving ecliptic, but are for the periodic and Poisson terms in two of the Euler angles (which they denote by ψ and ω) of the transformation from the fixed ecliptic of J2000 to the terrestrial equatorial frame of date. An unfortunate consequence of these differences is that direct comparison of their SMART97 nutation series with other series is not possible. One has to keep in mind, in addition, differences in signs between the different sets of variables: for example, $\omega_0 = -\varepsilon_0$ and $\Delta\omega$ corresponds to $-\Delta\varepsilon$ with the effects of ecliptic motion ignored.

3.10.6.2 Resolution of Elliptical Motions into Pairs of Circular Motions

With the coordinate frame chosen to be the GCRF in which the X- and Y-axes are in the directions of increasing longitude and latitude, respectively, the components Γ_1 and Γ_2 of the torque acting on an ellipsoidal Earth have the forms

$$\begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} = \begin{pmatrix} P_s \sin \Xi(t) \\ Q_c \cos \Xi(t) \end{pmatrix} \qquad [76]$$

with

$$\Xi(t) = (\omega_n t + \chi_\omega) \qquad [77]$$

where ω_n is the nutation frequency and χ_ω is the initial phase. The curve traced by the tip of this two-dimensional vector is circular if $Q_c = \pm P_s$, and elliptical otherwise. It should be kept in mind that Γ_1, Γ_2, P_s, Q_s (and other similar quantities introduced below) refer to a particular spectral component, though we have avoided adding an index ω to keep the notation simpler.

The Earth's response to the gravitational tidal forcing is not strictly in phase with the forcing, for a variety of reasons: Anelasticity of the mantle causes the tidal deformation of the solid Earth to lag behind; the ocean tide raised by the potential and hence the deformation produced by ocean loading are very much out of phase with the forcing and so on; and these effects affect the deformable Earth's rotational response to the gravitational torque. Consequently, the Cartesian components $X(t)$, $Y(t)$ of the nutation, for example, will not be in phase with the torque; $X(t)$ and $Y(t)$ will each have both cosine and sine terms:

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} X_s \sin \Xi + X_c \cos \Xi \\ Y_c \cos \Xi + Y_s \sin \Xi \end{pmatrix} \qquad [78]$$

X_s and Y_c belong to the part of the nutation that is 'in phase' with the torque, while X_c and Y_s pertain to the 'out-of-phase' part.

In most of the existing literature, the expansions into cosine and sine terms of various frequencies are written for $\Delta\psi$ and $\Delta\varepsilon$ rather than for X and Y, and the coefficients in those expansions are called the *coefficients of nutation*; thus, $X_s/\sin\varepsilon$ and Y_c are the in-phase coefficients of nutation, and $X_c/\sin\varepsilon$ and Y_s are the out-of-phase ones.

We can now resolve the two-vector with components (X, Y) into parts represented by the following two-vectors, each of which executes a prograde or retrograde circular motion:

$$\begin{pmatrix} \sin \Xi \\ -\cos \Xi \end{pmatrix}, \quad \begin{pmatrix} -\sin \Xi \\ -\cos \Xi \end{pmatrix}, \quad \begin{pmatrix} \cos \Xi \\ \sin \Xi \end{pmatrix}, \quad \begin{pmatrix} \cos \Xi \\ -\sin \Xi \end{pmatrix} \qquad [79]$$

A little reflection shows that if the $\Xi(t)$ is increasing with time, that is, if the frequency ω_n is positive, the first and third of the previously mentioned two-vectors rotate in the sense from the positive direction of the first axis of the equatorial reference frame toward that of the second axis. This is the same sense in

which the Earth rotates and is therefore said to be *prograde*. The second and fourth two-vectors rotate in the opposite or *retrograde* sense. The roles are reversed if ω_n is negative.

Expanding in terms of these prograde and retrograde vectors, we now have

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \eta^{+,\mathrm{ip}} \begin{pmatrix} \sin\Xi \\ -\cos\Xi \end{pmatrix} + \eta^{-,\mathrm{ip}} \begin{pmatrix} -\sin\Xi \\ -\cos\Xi \end{pmatrix} + \eta^{+,\mathrm{op}} \begin{pmatrix} \cos\Xi \\ \sin\Xi \end{pmatrix} + \eta^{-,\mathrm{op}} \begin{pmatrix} \cos\Xi \\ -\sin\Xi \end{pmatrix} \quad (80)$$

wherein the parts which come from the in-phase and out-of-phase parts are identified by the superscripts 'ip' and 'op,' respectively. The relations connecting the *amplitudes* $\eta^{\pm,\mathrm{ip}}$ of the in-phase parts of the nutation and $\eta^{\pm,\mathrm{op}}$ of the out-of-phase parts to the *coefficients of nutation* present in eqn [78] may be immediately written down. We have

$$\begin{aligned} \eta^{+,\mathrm{ip}} &= \frac{1}{2}(X_s - Y_c), \quad \eta^{-,\mathrm{ip}} = -\frac{1}{2}(X_s + Y_c) \\ \eta^{+,\mathrm{op}} &= \frac{1}{2}(X_c + Y_s), \quad \eta^{-,\mathrm{op}} = \frac{1}{2}(X_c - Y_s) \end{aligned} \quad [81]$$

Recall now that if $\omega_n > 0$, the prograde components of the motion are those with the coefficients $\eta^{+,\mathrm{ip}}$ and $\eta^{+,\mathrm{op}}$ and the retrograde ones have the coefficients $\eta^{-,\mathrm{ip}}$ and $\eta^{+,\mathrm{op}}$, while the connections are reversed if $\omega_n < 0$. So, if we identify the prograde and retrograde amplitudes by superscripts p and r, respectively, in the place of $\pm$, we see that for $\omega_n = \pm|\omega_n|$, $\eta^{p,\mathrm{ip}} = \eta^{\pm,\mathrm{ip}}$ and $\eta^{p,\mathrm{op}} = \eta^{\pm,\mathrm{op}}$, while $\eta^{r,\mathrm{ip}} = \eta^{\mp,\mathrm{ip}}$ and $\eta^{r,\mathrm{op}} = \eta^{\mp,\mathrm{op}}$. These relations may be written, with the use of eqn [81], as the following expressions for the prograde and retrograde amplitudes in terms of the coefficients of nutation:

$$\begin{aligned} \eta^{p,\mathrm{ip}} &= \frac{1}{2}\left(\frac{\omega_n}{|\omega_n|}X_s - Y_c\right) \\ \eta^{r,\mathrm{ip}} &= -\frac{1}{2}\left(\frac{\omega_n}{|\omega_n|}X_s - Y_c\right) \\ \eta^{p,\mathrm{op}} &= \frac{1}{2}\left(X_c + \frac{\omega_n}{|\omega_n|}Y_s\right) \\ \eta^{r,\mathrm{op}} &= \frac{1}{2}\left(X_c - \frac{\omega_n}{|\omega_n|}Y_s\right) \end{aligned} \quad [82]$$

Conversely,

$$\begin{aligned} X_s &= \frac{\omega_n}{|\omega_n|}(\eta^{p,\mathrm{ip}} - \eta^{r,\mathrm{ip}}), \quad Y_c = -(\eta^{p,\mathrm{ip}} + \eta^{r,\mathrm{ip}}) \\ X_c &= (\eta^{p,\mathrm{op}} + \eta^{r,\mathrm{op}}), \quad Y_s = \frac{\omega_n}{|\omega_n|}(\eta^{p,\mathrm{op}} - \eta^{r,\mathrm{op}}) \end{aligned} \quad [83]$$

In the literature, the relations in the preceding text are practically always written in terms of $\Delta\psi$ and $\Delta\varepsilon$ rather than in terms of X and Y. The transition from the relations obtained earlier is accomplished by the replacements

$$\begin{aligned} X_s &\to A\sin\varepsilon, \quad Y_c \to B \\ X_c &\to A''\sin\varepsilon, \quad Y_s \to B'' \end{aligned} \quad [84]$$

In these relations, ε is treated as a constant equal to ε_0 (i.e., neglecting the small increments to ε_0 from nutation in obliquity).

Complex combinations like $\tilde{\Gamma} = \Gamma_1 + i\Gamma_2$ and $X + iY$ are of considerable interest in the context of nutations. (We have already employed such combinations to advantage in Section 3.10.3.2.2.) It may be verified, by applying the developments presented earlier, that $X + iY$ may be written as

$$X + iY = \tilde{\eta}^p e^{i(\omega_n/|\omega_n|)\Xi} + \tilde{\eta}^r e^{-i(\omega_n/|\omega_n|)\Xi} \quad [85]$$

with

$$\tilde{\eta}^p = \eta^{p,\mathrm{ip}} + i\eta^{p,\mathrm{op}}, \quad \tilde{\eta}^r = \eta^{r,\mathrm{ip}} + i\eta^{r,\mathrm{op}} \quad [86]$$

$\tilde{\eta}^p$ and $\tilde{\eta}^r$ are the complex amplitudes of the prograde and retrograde components of the nutation with the in-phase coefficients X_s, Y_c and out-of-phase coefficients X_c, Y_s as in eqn [78].

It is to be noted that whichever term of eqn [85] has a positive value for the frequency (ω_n in the first term and $-\omega_n$ in the second one) that multiplies i t in the phase factor represents a prograde circular motion and the other, with a negative value of the frequency, represents retrograde motion. This is a fact worth keeping in mind for future reference.

3.10.6.3 Nutation Series

Up to this point, we have been focusing on one spectral component of the nutational motion. To represent the total motion, we have to add to the quantities appearing in the expressions on the right-hand side of eqn [78] an index n labeling the spectral terms of the various frequencies $\omega_n = d\Xi_n/dt$ present in the nutation and then sum over all the terms. The resulting series provide the spectral representation of $X(t)$ and $Y(t)$. The series for $\Delta\psi$ is $1/\sin\varepsilon$ times that of $X(t)$, while the series for $\Delta\varepsilon$ is the same as that of $Y(t)$.

However, the series that are thus obtained are not quite complete. A reason is that the spectral components of the torque are not strictly harmonic as would appear from the expressions in eqn [76]; the amplitudes have a slow variation with time, which may be taken to be linear. Moreover, the fundamental arguments of nutation (and hence the argument $\Xi(t)$, which appears in the torque terms) are not strictly linear in time, which is to say that the nutation frequencies also are varying, though very slowly. The combined effect of the slow variations of the amplitudes and the arguments is that the amplitudes of the spectral terms in the nutation series are not constants. The time dependence of the amplitudes is adequately taken into account in the classical nutation series for $\Delta\psi$ and $\Delta\varepsilon$ through a linear dependence on time. With this enhancement, the series referred to in the last paragraph take the following form:

$$\begin{aligned} \Delta\psi(t) &= \sum_n \left[(A_n + A_n' t)\sin\Xi_n(t) + A_n'' \cos\Xi_n(t)\right] \\ \Delta\varepsilon(t) &= \sum_n \left[(B_n + B_n' t)\cos\Xi_n(t) + B_n'' \sin\Xi_n(t)\right] \end{aligned} \quad [87]$$

Nutation series have the same general form as in the preceding text irrespective of whether they are for the figure axis, rotation axis, or the angular momentum axis, but the coefficients of nutation differ from one to the other.

Nutations have periods extending from thousands of years down to fractions of a day, as we have seen earlier. The largest of them are the 18.6-year, 9.3-year, annual, semiannual, and 13.66-day nutations. The in-phase coefficients A_n and B_n of the nutations of the figure axis are shown in Tables 1 and 2 for these five periods. The values are from recent computations

Table 1 Coefficients in longitude and obliquity of prominent nutations (in mas)

Period[a]	Rigid Earth		Nonrigid Earth	
	Longitude	Obliquity	Longitude	Obliquity
−18.6 years	−17 281	9228	−17 208	9205
−9.3 years	209	−90	207	−90
1 year	126	0	148	7
0.5 years	−1277	553	−1317	573
13.66 days	−222	95	−228	98

[a]Some of the periods appear with negative signs. They have been historically assigned to those nutations in which the retrograde component dominates over the prograde one (compare **Table 2**).

Table 2 Amplitudes of prograde and retrograde circular nutations (in mas)

Period[a]	Rigid Earth		Nonrigid Earth	
	Retrograde	Prograde	Retrograde	Prograde
18.6 years	−8051	−1177	−8025	−1180
9.3 years	87	4	86	4
1 year	−25	25	−33	26
0.5 year	−23	−531	−25	−548
13.66 days	−3	−92	−4	−94

[a]The periods shown are for the prograde nutations. The retrograde nutations have the corresponding negative values for their periods.

and are given both for a nonrigid Earth model and for a hypothetical rigid Earth. The entries in the table bring out the fact that the nutations of the figure axis are significantly affected by nonrigidity of the Earth, especially at some frequencies; in particular, the amplitude of the retrograde annual nutation of the nonrigid Earth is over 30% higher than that of the rigid Earth, while that of the prograde annual nutation is hardly affected by nonrigidity. The reason for the differing levels of sensitivity of different spectral components is that certain normal modes and associated resonances (especially the FCN) exist and affect the Earth's response to the torque in a frequency-dependent fashion in the case of the nonrigid Earth but not in the rigid case, as will be seen from the theory presented in later sections.

3.10.7 Kinematic Relations between the Nutation of the Figure Axis and the Wobble

Consider the motion of the Earth's figure axis relative to the equatorial CRF of J2000 (see **Section 3.10.2.2**). Recall that the coordinates of the pole of this axis in this frame at an instant t are $X(t)$, $Y(t)$. The Z coordinate of the pole is of no interest here.

The motion of the pole in the interval $(t, t+\mathrm{d}t)$ changes its coordinates by $(\mathrm{d}X, \mathrm{d}Y, 0)$. We transform this infinitesimal vector to the TRF now, in order to establish its relation to the components of the wobble. We seek to obtain the transformed components only to the first order in small quantities. For this limited purpose, we may take the transformation matrix to the zeroth order, meaning that we ignore the first-order offsets between the Z-axis of the celestial frame, on the one hand, and the figure and rotation axes of the Earth, on the other, and also ignore the variations in the spin rate. In this approximation, the equators of the celestial and terrestrial frames coincide, and the transformation from the former to the latter is simply a uniform rotation with angular velocity Ω_0 about the Z-axis of the celestial frame. So the components of the infinitesimal displacement vector of the figure axis pole in the terrestrial frame are

$$\begin{aligned} \mathrm{d}x &= \mathrm{d}X\cos\Omega_0(t-t_0) + \mathrm{d}Y\sin\Omega_0(t-t_0) \\ \mathrm{d}y &= -\mathrm{d}X\sin\Omega_0(t-t_0) + \mathrm{d}Y\cos\Omega_0(t-t_0) \end{aligned} \quad [88]$$

where t_0 is the instant of coincidence of the first axes of the two frames. The errors in these expressions are of the second order in the neglected small quantities.

The displacement $(\mathrm{d}x, \mathrm{d}y)$ of the figure axis pole in the terrestrial frame is, however, determined directly by the angular velocity vector. The Earth's rotation with angular velocity $\boldsymbol{\Omega}$ has the effect that any mantle-fixed vector $\mathbf{V}$ gets displaced by an amount $(\boldsymbol{\Omega}\times\mathbf{V})\mathrm{d}t$ in an infinitesimal time interval $\mathrm{d}t$. In particular, the displacement of the z-axis during the interval $(t,\ t+\mathrm{d}t)$ is $(\boldsymbol{\Omega}\times\hat{\mathbf{Z}})\,\mathrm{d}t$, with components $(\Omega_y,\ -\Omega_x,\ 0)\,\mathrm{d}t = \Omega_0(m_y,\ -m_x,\ 0)\mathrm{d}t$ relative to the terrestrial frame of the instant t. Since the z-axis is the figure axis, we have thus

$$\mathrm{d}x = \Omega_0 m_y\,\mathrm{d}t, \quad \mathrm{d}y = -\Omega_0 m_x\,\mathrm{d}t \quad [89]$$

We obtain the relations that we seek on equating the expressions [88] and [89]:

$$\begin{aligned} \Omega_0 m_x &= \dot{X}\sin\Omega_0(t-t_0) - \dot{Y}\cos\Omega_0(t-t_0) \\ \Omega_0 m_y &= \dot{X}\cos\Omega_0(t-t_0) - \dot{Y}\sin\Omega_0(t-t_0) \end{aligned} \quad [90]$$

The familiar complex combination of this pair of equations now yields

$$\begin{aligned} \Omega_0\tilde{m} &= i\left(\dot{X}+i\dot{Y}\right)\mathrm{e}^{-i\Omega_0(t-t_0)} \\ &= i\frac{\mathrm{d}\tilde{\eta}}{\mathrm{d}t}\mathrm{e}^{-i\Omega_0(t-t_0)} \end{aligned} \quad [91]$$

where

$$\tilde{\eta} = X + iY \quad [92]$$

It is important to note that no dynamical considerations have been invoked while deriving the result [91]. Therefore, it is valid for rigid and nonrigid Earth models alike. Since it is of purely kinematic origin, it is commonly referred to as the *kinematic relation*.

3.10.7.1 Kinematic Relations in the Frequency Domain

For a circular nutation of some frequency ω_n (using the subscript n here for nutation),

$$\begin{aligned} X &= \tilde{\eta}(\omega_\mathrm{n})\cos(\omega_\mathrm{n}t+\chi_\mathrm{n}) \\ Y &= \tilde{\eta}(\omega_\mathrm{n})\sin(\omega_\mathrm{n}t+\chi_\mathrm{n}) \\ X+iY &= \tilde{\eta} = \tilde{\eta}(\omega_\mathrm{n})\mathrm{e}^{i(\omega_\mathrm{n}t+\chi_\mathrm{n})} \end{aligned} \quad [93]$$

where χ_n is a phase that depends on the frequency ω_n. On introducing this expression into the first of the equalities in eqn [91], it becomes obvious that the form of $\tilde{m}$ is given by

$$\tilde{m} = \tilde{m}(\omega_w)e^{i(\omega_w t+\chi_w)} \quad [94]$$

with

$$\tilde{m}(\omega_w) = -\frac{\omega_n}{\Omega_0}\tilde{\eta}(\omega_n), \quad \omega_w = \omega_n - \Omega_0 \quad [95]$$

(The subscript w labels quantities pertaining to the wobble.) These equations constitute the kinematic relations in the spectral domain; they correspond to the kinematic relation [91] in the time domain.

An additional relation, which is not usually written down explicitly, is obtained on considering the zero-frequency part of $\dot{X}$, which is $\dot{\psi}_A \sin\varepsilon$ where $\dot{\psi}_A$ is the rate of precession in longitude. It is evident from eqn [91] that the corresponding spectral component of $\tilde{m}$ has frequency $(-\Omega_0)$. It follows then that

$$\Omega_0\tilde{m}(-\Omega_0) = i\dot{\psi}_A \sin\varepsilon - \dot{\varepsilon}_A \quad [96]$$

With ε approximated by ε_0, the previous equation expresses the precession rate $\dot{\psi}_A$ in terms of the wobble amplitude at the frequency $-\Omega_0$ (-1 cpsd) as $-i(\Omega_0/\sin\varepsilon_0)\tilde{m}(-\Omega_0)$.

Another relation that is of importance concerns polar motion. We recall from Section 3.10.2.5.1 that the motion of the CIP in the celestial frame involves only frequencies in the low-frequency bands. This means that for all other frequencies, the CIP coincides with the pole of the celestial frame. But these are the frequencies, after transforming to the terrestrial frame, that are contained in polar motion. It follows that polar motion is effectively the motion of the pole of the celestial frame relative to the terrestrial frame. But the reverse of this motion, that is, the motion of the pole of the terrestrial frame (which is the pole of the figure axis) in the celestial frame, is the nutation $\tilde{\eta}$ of the figure axis. Therefore, the spectral component of frequency ω_p of the polar motion as defined by the IAU conventions may be obtained simply by transforming the negative of the corresponding spectral component of the nutation of the figure axis from eqn [93] to the terrestrial frame:

$$\begin{aligned}\tilde{p}_{\omega_p}(t) &= -\tilde{\eta}(\omega_n)e^{i(\omega_n t+\chi_n)}e^{-i\Omega_0 t}\\ \omega_p &= \omega_w = \omega_n - \Omega_0\end{aligned} \quad [97]$$

On using the kinematic relation [95] between $\tilde{\eta}$ and $\tilde{m}$, we finally obtain the relation between the amplitudes of the polar motion and the corresponding wobble:

$$\tilde{p}(\omega_p) = -\tilde{\eta}(\omega_n) = \frac{\Omega_0}{\omega_n}\tilde{m}(\omega_w) \quad [98]$$

Clearly, a polar motion of the CIP, in the frequency domain in which it is defined, is not identical to the wobble of the same frequency.

3.10.7.2 Implications of the Kinematic Relations

Now, we examine briefly the implications of the kinematic relations.

First and foremost is that these relations provide a formal proof of the existence of a circular nutation associated with every circular wobble and *vice versa*; the frequency of the nutation is algebraically greater than that of the associated wobble by 1 cpsd. This important general relation is simply a consequence of the Earth's diurnal rotation in space with angular velocity Ω_0. This relationship was already seen in a qualitative way earlier: it was noted in Section 3.10.4.2 that the spectrum of the torque in the terrestrial frame is in the retrograde diurnal band, and it was shown explicitly in Section 3.10.5.2 that the spectrum in the celestial frame is in the low-frequency band.

The amplitude relation in eqn [95] shows that the amplitude $\tilde{\eta}(\omega_n)$ of the nutation contains a factor $1/\omega_n$, which simulates a normal mode at $\omega_n = 0$. Unlike a proper normal mode, it has no connection to any property of the rotating Earth. It has been referred to in the literature as the 'tilt-over mode,' following Smith (1974) and Wahr (1981). The resonance factor $1/\omega_n$ produces a great enhancement of the nutation amplitude compared with the associated wobble amplitude when the frequency is very small, their ratio $(-\Omega_0/\omega_n)$ being, apart from a minus sign, the period of the nutation in sidereal days. Thus, the amplitude of the retrograde 18.6-year nutation (period $= -6798$ solar days $= -6816$ sidereal days) is 6816 times as large as that of the corresponding retrograde diurnal wobble. There is no such resonant kinematic factor in other frequency bands, and this is a major reason for the overriding importance of the low-frequency band in nutations. As far as other normal modes are concerned, each wobble normal mode and the resonance in the wobble amplitude that is associated with it have their counterparts in the nutation amplitude, of course with a shift in frequencies by 1 cpsd.

As an immediate application of the kinematic relations, we use the amplitude relation in eqn [95] to arrive at the amplitude $\tilde{\eta}(\omega_n)$ of the nutation of the *rigid Earth* that is produced by a torque $\tilde{\Gamma}(\sigma)$ having frequency σ cpsd in the terrestrial frame. Introducing the wobble amplitude eqn [19] of the rigid Earth into this relation, we obtain

$$\tilde{\eta}_R(\omega_n) = i\frac{\tilde{\Gamma}(\omega_w)}{A(\omega_w - \Omega_0 e)(\Omega_0 + \omega_w)} \quad [99]$$

As for the precession rate, one finds from eqn [19] at the particular frequency $\omega_w = -\Omega_0$, together with eqn [96], that

$$\dot{\psi}_A = \frac{\tilde{\Gamma}(-\Omega_0)}{A\Omega_0^2(1+e)} \quad [100]$$

3.10.7.3 Transfer Function

Consider two Earth models, one rigid and the other nonrigid, both of which are forced by the same torque $\tilde{\Gamma}(\omega_w)$. (In the simplest case when both are axially symmetrical and ellipsoidal, this means that both must have the same value for the dynamical ellipticity e.) Let the wobbles of the rigid and nonrigid Earth models be $\tilde{m}_R(\omega_w)$ and $\tilde{m}(\omega_w)$, respectively. The former may be taken from $\tilde{m}$ of eqn [19] of Section 3.10.3.2.2, which is for the rigid Earth. Both $\tilde{m}_R$ and $\tilde{m}$ are proportional to $\tilde{\Gamma}(\omega_w)$, and therefore, the ratio $\tilde{m}(\omega_w)/\tilde{m}_R(\omega_w)$ is independent of the torque. Moreover, as was observed in the last section, the kinematic relation [91] and its frequency domain version [95] are valid for any type of Earth model, rigid or nonrigid. Therefore, it follows from the latter equation that

$$T_{\mathrm{n}}(\omega_{\mathrm{n}}) \equiv \frac{\tilde{\eta}(\omega_{\mathrm{n}})}{\tilde{\eta}_{\mathrm{R}}(\omega_{\mathrm{n}})} = \frac{\tilde{m}(\omega_{\mathrm{w}})}{\tilde{m}_{\mathrm{R}}(\omega_{\mathrm{w}})} \equiv T_{\mathrm{w}}(\omega_{\mathrm{w}}) \qquad [101]$$

These ratios, which are independent of the forcing torque as noted earlier, are the *transfer functions* for nutation and wobble; the two are equal. It is useful to write these relations also with the alternative notation for frequencies:

$$T_{\mathrm{n}}(\nu) \equiv \frac{\tilde{\eta}(\nu)}{\tilde{\eta}_{\mathrm{R}}(\nu)} = \frac{\tilde{m}(\sigma)}{\tilde{m}_{\mathrm{R}}(\sigma)} \equiv T_{\mathrm{w}}(\sigma) \qquad [102]$$

where ν and σ are the frequencies of the nutation and the wobble, respectively, expressed in cpsd:

$$\nu = \omega_{\mathrm{n}}/\Omega_0, \quad \sigma = \omega_{\mathrm{w}}/\Omega_0, \quad \nu = \sigma + 1 \qquad [103]$$

the last equality being one of the kinematic relations.

The transfer functions are of great utility for the following reason. If one solves for wobble amplitude $\tilde{m}(\sigma)$ of the Earth model of interest by solving the relevant equations of motion with inputs from the Earth's structure pertaining to that model, one can compute the transfer function $T_{\mathrm{w}}(\sigma)$ (which is independent of the amplitude of the forcing), by dividing the amplitude by $\tilde{m}_{\mathrm{R}}(\sigma)$ as given by eqn [19] or [55]. Since $T_{\mathrm{n}}(\nu) = T_{\mathrm{w}}(\sigma)$, we can immediately obtain the nutation amplitudes of the chosen Earth model as

$$\tilde{\eta}(\nu) = T_{\mathrm{w}}(\sigma)\tilde{\eta}_{\mathrm{R}}(\nu) = -T_{\mathrm{w}}(\sigma)\tilde{m}_{\mathrm{R}}(\sigma)/(1+\sigma) \qquad [104]$$

where the last step makes use of the kinematic relations. Of course, $\tilde{m}_{\mathrm{R}}(\sigma)$ may be readily calculated using the tide tables, as explained in eqn [55].

Alternatively, $\tilde{n}_{\mathrm{R}}(\nu)$ in the last equation may be taken directly from tables of rigid Earth nutation amplitudes. Nutations of the rigid Earth have been studied extensively, and highly accurate results are available for the rigid Earth nutation amplitudes. The important point is that the computation of the torques exerted by the solar system bodies, starting with their time-dependent positions as given by the ephemerides, needs to be done only once, at the stage of working out the rigid Earth amplitudes. The use of the transfer functions enables one to avoid repeating this exercise for each nonrigid Earth model.

3.10.7.4 Relations Connecting the Nutations of Different Axes: Oppolzer Terms

The nutation of the rotation axis can be related to that of the figure axis with the aid of the kinematic relation [91], as we shall now show. This relation, like the kinematic relation, holds good whether the Earth is rigid or nonrigid.

By definition, the temporal variations of the vectors from the pole of the Z-axis of the celestial frame to the poles of the figure axis and the rotation axis, as seen from the celestial frame, constitute the nutations of these two axes. We denote the complex combinations of the (equatorial) components of the pole of the figure axis in the celestial frame by $(X+iY)$ and that of the rotation axis by $(X_{\mathrm{R}}+iY_{\mathrm{R}})$. (Caution: In this section, R does not stand for 'Rigid' as elsewhere, but for 'rotation.') Now, it is evident that the difference $(X_{\mathrm{R}}+iY_{\mathrm{R}})-(X+iY)$ gives the complex representation of the components (in the celestial frame) of the vector from the figure axis pole to the rotation pole. It is the variation of this vector, as seen in the *terrestrial* frame, that constitutes the wobble, represented by m; its components in the celestial frame will then be represented by $\tilde{m}e^{i\Omega_0(t-t_0)}$, under the same approximations as in the second paragraph of Section 3.10.7. Consequently, we have

$$(X_R + iY_R) - (X + iY) = \tilde{m}e^{i\Omega_0(t-t_0)} \qquad [105]$$

We can now use the kinematic relation [91] to eliminate $\tilde{m}$ from the previous equation. The result is the relation that we seek:

$$X_R + iY_R = (X + iY) + \frac{i}{\Omega_0}\left(\dot{X} + i\dot{Y}\right) \qquad [106]$$

For a spectral component of the nutation with frequency ω_{n}, one writes

$$\begin{aligned} X + iY &= \tilde{\eta}(\omega_{\mathrm{n}})e^{i(\omega_{\mathrm{n}}t+\chi_{\mathrm{n}})} \\ X_{\mathrm{R}} + iY_{\mathrm{R}} &= \tilde{\eta}_{\mathrm{R}}(\omega_{\mathrm{n}})e^{i(\omega_{\mathrm{n}}t+\chi_{\mathrm{n}})} \end{aligned} \qquad [107]$$

and the foregoing relation becomes

$$\tilde{\eta}_{\mathrm{R}}(\omega_{\mathrm{n}}) = \left(1 - \frac{\omega_{\mathrm{n}}}{\Omega_0}\right)\tilde{\eta}(\omega_{\mathrm{n}}) \qquad [108]$$

We see that the fractional difference between the nutation amplitudes of the rotation axis and the figure axis is the frequency of the nutation expressed in cpsd.

It is of interest to note that the precession of the figure axis in longitude, which is represented by $\dot{\psi}_{\mathrm{A}}t$, leads to the precession part $\dot{\psi}_{\mathrm{A}}(t + i/\Omega_0)$ in $(X_{\mathrm{R}}+iY_{\mathrm{R}})$. Thus, one finds that the rotation pole has, expectedly, the same rate of precession in longitude as the pole of the figure axis, but the rotation pole is offset by a constant amount $\left(\dot{\psi}_{\mathrm{A}}/\Omega_0\right)$ in obliquity.

One may ask whether the nutation of the angular momentum axis is also related in a simple manner to the nutations of the other two axes. The answer is in the negative, in general. The reason is that the relation of the angular momentum vector to the rotation vector depends very much on the structure of the Earth, and so the relation between the nutations of the angular momentum and rotation axes differs from one Earth model to another. The particular case of a rigid Earth is of considerable interest, however, because classical treatments of the rotation variations of the rigid Earth have had, as their primary output, the nutations of the angular momentum axis; the nutations of the figure and rotation axes were then inferred. The relations used for this last step will now be derived, taking the Earth to be a rigid axially symmetrical ellipsoid. The components of the angular momentum vector in a principal axis frame are then $H_x = A\Omega_0 m_x$, $H_y = A\Omega_0 m_y$, $H_z = C\Omega_0(1+m_z)$, and the equatorial components of the unit vector along $\mathbf{H}$ are $(A/C)\ m_x$, $(A/C)m_y$ to the lowest order in m_x, m_y, m_z. These are the components of the vector offset of the pole of the angular momentum axis from that of the figure axis (the z-axis); they are simply (A/C) times the corresponding components (m_x, m_y) for the pole of the rotation axis. The scalar factor (A/C) is of course independent of whether the offset vectors of the angular momentum axis and the rotation axis are viewed from the terrestrial or the celestial frame. Therefore, we have

$$(X_H + iY_H) - (X + iY) = (A/C)[(X_{\mathrm{R}} + iY_{\mathrm{R}}) - (X + iY)] \qquad [109]$$

This means that when considering the offset of the angular momentum axis, a relation of the form [105] holds with $\tilde{m}$ replaced by $(A/C)\tilde{m} = \tilde{m}/(1+e)$. Consequently, we have, instead of eqn [106], the relation

$$X_H + iY_H = (X + iY) + \frac{i}{(1+e)\Omega_0}\left(\dot{X} + i\dot{Y}\right) \qquad [110]$$

The corresponding relation for the amplitudes of the respective spectral components of frequency ω_n is

$$\tilde{\eta}_H(\omega_n) = \left(1 - \frac{\omega_n}{(1+e)\Omega_0}\right)\tilde{\eta}(\omega_n) \qquad [111]$$

Therefore,

$$\tilde{\eta}(\omega_n) - \tilde{\eta}_H(\omega_n) = \frac{\omega_n}{(1+e)\Omega_0 - \omega_n}\tilde{\eta}_H \qquad [112]$$

This is called the *Oppolzer term for the figure axis*. It enables one to compute the amplitude of the nutation of the figure axis, given that of the angular momentum axis. The *Oppolzer term for the rotation axis* can be obtained by combining the previous equation with eqn [108]. The result is that

$$\tilde{\eta}_R(\omega_n) - \tilde{\eta}_H(\omega_n) = -\frac{e\omega_n}{(1+e)\Omega_0 - \omega_n}\tilde{\eta}_H \qquad [113]$$

It must be kept in mind that the expressions for the Oppolzer terms [112] and [113] are valid only for the rigid Earth model considered, unlike the kinematic relation and the relation [108] between $\tilde{\eta}_R$ and $\tilde{\eta}$, which are independent of the Earth model.

Traditionally, nutation tables for the rigid Earth have been presented as tables of the coefficients in the spectral expansions of $\Delta\psi_H(t)$ and $\Delta\varepsilon_H(t)$ in terms of real simple harmonic (cosine and sine) functions of time, together with the corresponding coefficients for the Oppolzer terms. In the case of nonrigid Earth models, tables are presented, as a rule, for the nutations of the figure axis.

3.10.8 Rigid Earth Nutation

By 'rigid Earth,' we mean a rigid Earth model having the same density function $\rho(\mathbf{r})$ as for the actual Earth. It has therefore the same principal moments of inertia, and other moments of higher degree of the matter distribution, as the actual Earth. These parameters are determined from the spatial structure of the Earth's own external gravitational field, inferred from its influence on the orbits of low-flying Earth satellites observed by satellite laser ranging. The determination of the rotation variations should then, in principle, be a simple matter, if the time dependence of the gravitational potentials of the Moon, Sun, and the planets at the Earth is known. This dependence is determined by the orbital motions of the Moon around the Earth and of the Earth and other solar system bodies around the Sun. The orbits deviate from simple Kepler's ellipses for a variety of reasons: (1) perturbation of the lunar orbit around the Earth by the attraction of the Moon by the Sun and of the Earth's orbit around the Sun by the Moon's gravitation (the so-called three-body effects), (2) additional perturbations of these orbits by the planets (resulting in the 'indirect planetary effect' on Earth rotation) and also the so-called planetary-tilt effect on the lunar orbit, and (3) the 'J_2-tilt effect' due to the perturbation of the Moon's orbit by the noncentral part of the Earth's gravitational potential that arises from the Earth's ellipticity. All these orbit perturbations are reflected in the lunar and solar gravitational potentials at the Earth and hence in their torques on the Earth. Additional torques arise from the direct action of the planets on the ellipticity of the Earth, referred to as the 'direct planetary effect.' The ELP2000 ephemerides provide the time-dependent position of the Moon in analytic form, and VS0P2000 presents similar information for the Sun and the planets. These ephemerides are recent versions of the ELP and VSOP series of ephemerides beginning with Chapront-Touzé and Chapront (1983) and Bretagnon and Francou (1988), respectively. The ephemerides of the Sun and the Moon include the perturbations mentioned in the preceding text. Treatments of the rotation variations of the rigid Earth fall into two broad classes: those based on the Hamiltonian formulation of mechanics and those making use of the torque equation governing rotational motion.

The Hamiltonian approach begins generally with a treatment of the 'main problem' where the influences of the Moon and Sun on the Earth's rotation are considered, taking into account the Sun's perturbation of the lunar orbit and the Moon's perturbation of the Earth–Sun orbital motion and, in recent treatments, also the planetary perturbations of both these orbits. The direct planetary torques on the Earth are treated separately, as are other second-order effects like the 'crossed-nutation coupling,' a term that refers to the change in the torque on the Earth due to the changes in the orientation of the Earth itself (due to precession–nutation) and the consequent increments to these motions.

The torque approach makes direct use of the ephemerides, which already take account of all the orbit perturbations. In either approach, the Earth parameter, which plays the dominant role in the gravitational torque on the Earth, is the dynamical ellipticity, but triaxiality and the higher-degree geopotential coefficients that play lesser roles are also considered. In addition, second-order increments to the gravitational torque exerted by the Moon and Sun due to changes in Earth's orientation caused by nutation and precession are also taken into account in recent rigid Earth nutation theories.

3.10.8.1 Hamiltonian Approach

In the Hamiltonian approach, the starting point is the Hamiltonian K of the rotating Earth. (The standard notation for the Hamiltonian is H; it has been changed here because of the use of H for another quantity in the present context.) K is the sum of the Earth's rotational kinetic energy and the potential energy of its gravitational interaction with the Moon and Sun. (The relatively small interactions with the planets are dealt with separately.) Recent computations of the Earth's precession and nutation by Souchay et al. (1999) are the culmination of developments starting from the work of Kinoshita (1977) based on the Hamiltonian approach. The presentation in the succeeding text is a very brief overview of that work.

The Hamiltonian of a system is a function of a set of coordinate variables q_i and corresponding momentum variables p_i, which, taken together, describe the state of the system at any given time; (q_i, p_i) for any given i is said to constitute a

canonically conjugate pair. Given the expression $K(q, p)$ for the Hamiltonian, the equations of motion are given by

$$\dot{q}_i = \frac{\partial K}{\partial p_i}, \quad \dot{p}_i = \frac{\partial K}{\partial q_i} \qquad [114]$$

for all i.

The Hamiltonian for the rotational motion of the Earth is

$$K = \sum_i \frac{L_x^2}{2A} + \frac{L_y^2}{2B} + \frac{L_z^2}{2C} + M_B W_E(r_B, \theta_B, \lambda_B) \qquad [115]$$

wherein the kinetic energy involves the components of the angular momentum vector in the terrestrial frame (which are denoted here by L_x, L_y, L_z in a departure from earlier notation) and the potential energy is expressed through the potential of the Earth at the position of the body B. (To be specific, we take the body to be the Moon; the case of the Sun is entirely similar). Since θ_B and λ_B characterize the direction of the Moon in the TRF, they depend on the orientation of this frame in space, that is, in relation to an inertial reference frame. This orientation may be represented by the transformation that carries the latter frame to the former.

It is customary, in Hamiltonian treatments of the rotation variations of the rigid Earth, to place the nutation and precession of its angular momentum axis in the foreground and to infer from them subsequently the motions of the figure and rotation axes. Consequently, the inertial plane (which is the principal plane normal to the third axis of the inertial frame), the 'angular momentum plane' (or Andoyer plane) normal to the angular momentum vector, and the principal plane of the terrestrial frame (the equatorial plane) are all made use of in the transformation from the inertial frame to the terrestrial one (see **Figure 8** for the definition of the angles used). The transformation consists of (1) a rotation through an angle h about the third axis of the inertial frame to bring its first axis into a new position along the nodal line (intersection) of the inertial and angular momentum planes, (2) a rotation through an angle/about this new axis to carry the third axis of the inertial frame into the direction of the angular momentum vector, (3) a rotation about this vector through an angle g to take the first axis from the earlier mentioned nodal line to the node of the angular momentum plane and the equatorial plane, (4) a rotation through angle I about this node to carry the third axis over from the angular momentum axis to the third axis of the terrestrial frame, and (5) a rotation through an angle l about the last-mentioned axis to bring the first axis into alignment with the first axis (at 0 longitude) of the terrestrial frame. The rotation angles h, g, l are taken as the coordinate variables q_i ($i = 1, 2, 3$), describing Earth orientation, and the angular momentum components H, G, L along the axes about which these rotations are made are taken as the conjugate momentum variables p_i for the purpose of the Hamiltonian formulation. (H and L are evidently the components of the angular momentum vector along the Z-axis of the inertial frame and the z-axis of the terrestrial frame, respectively, and G is the magnitude of the angular momentum vector.) These three pairs of canonically conjugate variables are called the Andoyer variables (Andoyer, 1923). The two other quantities I and J are not independent of these: $\cos I = H/G$, $\cos J = L/G$.

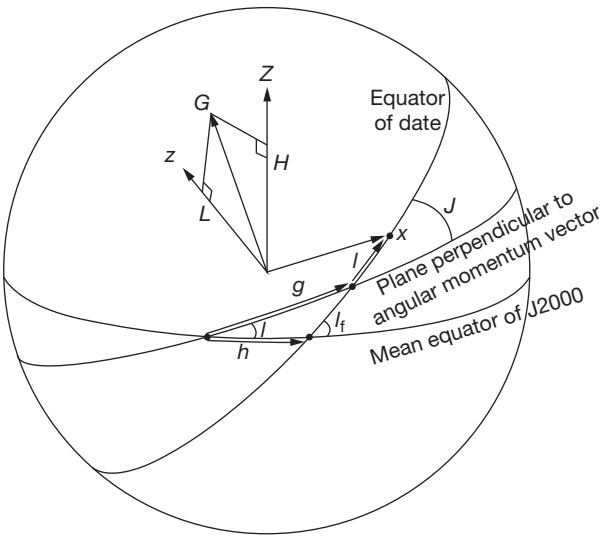


Figure 8 Angles used in the transformations between the celestial and terrestrial frames in the Hamiltonian approach for rigid Earth nutation computation.

The nutation and precession of the angular momentum axis in longitude consist of the motions of this axis around the third axis of the inertial plane. This is precisely what variations Δh in the angle h describe. The nutation–precession in obliquity is represented by the variations ΔI in the inclination of the angular momentum axis to the inertial axis; this quantity may be expressed in terms of variations in the Andoyer variables H and G: $\Delta I = [(H/G)\Delta G - \Delta H]/(G \sin I)$. They may be determined, therefore, by solving the Hamiltonian equations for the Andoyer variables.

To set up the Hamiltonian equations, it is necessary to express the total energy given by eqn [115] as a function of the Andoyer variables representing Earth rotation, on the one hand, and the position coordinates of the Moon with reference to the inertial frame, on the other. The orbital motion of the Moon makes its coordinates time-dependent. The ephemerides (ELP2000 in the case of the Moon) express the time dependence of each of the position coordinates as a spectral expansion in which the argument of each term is an integer linear combination of the Delaunay fundamental arguments and additional arguments referring to the motions of the planets (see **Section 3.10.6.1**); the planetary arguments enter because of the gravitational perturbations they produce on the lunar orbit. Expansions of the same kind are also provided for the functions of the coordinates, which appear in the spherical harmonic expansion of the potential in the celestial frame. With the time dependence of the potential from the orbital motions of the solar system bodies thus made explicit, the Hamiltonian equations $\dot{h} = \partial K/\partial H$, $\dot{H} = -\partial K/\partial h$ and so on yield the three pairs of time derivatives as functions of the six Andoyer variables and of the time variable appearing through the spectral expansions.

Kinoshita (1977) applied a method due to Hori (1966) to facilitate the solution of these equations by means of a transformation of the Andoyer variables and their functions, which preserves the form of Hamilton's equations while

expressing them in terms of the new (transformed) variables. The transformation, which is carried out perturbatively, is designed to cause the new Hamiltonian to have only a secular variation with time, while all periodicities enter through the 'determining function' through which the transformation is accomplished. A detailed treatment of the transformation and the subsequent solution of the equations are beyond our scope. The interested reader may refer to Kinoshita (1977). For the inclusion of planetary influences, see Souchay and Kinoshita (1996, 1997). Various second-order effects are treated in later papers of Souchay, Kinoshita, and collaborators (Folgueira et al., 1998a,b; Kinoshita and Souchay, 1990; Souchay and Kinoshita, 1996, 1997), and the complete results are presented by Souchay et al. (1999) as the REN-2000 nutation series. This is the rigid Earth series used by Mathews et al. (2002) for convolution with their transfer function in constructing the MHB2000 series on which the IAU2000A nutation series is based.

3.10.8.2 Torque Approach

This approach is based on the solution of the torque equation either in the celestial frame or in the terrestrial frame, introduced in **Sections 3.10.3.1** and **3.10.3.2**, respectively.

3.10.8.2.1 Axially symmetrical Earth: simplified treatment

The Earth's structure parameters (the principal moments of inertia and geopotential coefficients of higher orders) are constant in the terrestrial frame, and therefore, the torque on the Earth is readily determined in this frame, as noted earlier. Transformation of the torque to the celestial frame is easy only if the Earth is taken to be axially symmetrical and the effects of the Earth's nutation, precession, and wobble on the torque are ignored. We have seen in **Section 3.10.4.5** that the structure parameters that appear in the torque on an axially symmetrical Earth are the J_n (equal to $-C_{n0}$) and that the torque proportional to J_n are given by eqn [61] in the terrestrial frame. Under the approximations stated earlier, the corresponding torque in the celestial frame may be obtained simply by replacing θ_B and λ_B by $(\pi/2 - \delta_B)$ and α_B, respectively, where δ_B is the declination and α_B the right ascension of the body B in the equatorial CRF. The reasoning is the same as in the special case $n = 2$ considered in **Section 3.10.5.1** and will not be repeated here. The relations between the Cartesian coordinates used there and the polar coordinates δ and α are $X_B = r_B \cos\delta_B \cos\alpha_B$, $Y_B = r_B \cos\delta_B \sin\alpha_B$, $Z_B = r_B \sin\delta_B$.

The time dependence of the torque may now be made explicit by using the ELP or VSOP ephemerides, as the case may be. The nutation of the angular momentum axis is then trivially obtained by integration of the torque (see **Section 3.10.3.1**). This approach has been used by Roosbeek and Dehant (1998); they identified the corrections that need to be applied to the nutation series obtained directly from the integration. Their final results constitute the RDAN (Roosbeek-Dehant Analytical Nutation) nutation series. Elegant analytic derivations have been provided by Williams (1994) for the contributions to nutation and precession from various small effects.

3.10.8.2.2 Rigorous treatment of the general case

The transformation of the torque equations taking account of the full range of geopotential coefficients in the terrestrial frame (ITRF) to equivalent equations in the fixed ecliptic celestial reference frame (ECRF) of J2000 can be carried out rigorously, as was done by Bretagnon et al. (1997, 1998). (We denote this celestial frame by ECRF0, with 0 serving as a reminder that it is based on a fixed ecliptic.) The transformation is effected through a sequence of three rotations. The angles of the rotations, which these authors represent by ψ, ω, and φ, are the Euler angles of the transformation. The first two rotations bring one over from the ecliptic plane to the equatorial plane while taking the first axis of ECRF0 over into the nodal line of the two planes; the rotation through φ is about the third axis of the ITRF and takes the first axis from the nodal line to the first axis of the ITRF. If one desires that the final frame be the principal axis frame rather than the ITRF, one needs only to replace φ by $\tilde{\varphi} = \varphi + \alpha$, where α is the longitude (in the ITRF) of the equatorial principal axis of the least moment of inertia A.

The variations of φ (or of $\tilde{\varphi}$) are closely related to LOD variations. The variations of ψ and $(-\omega)$ as a result of the variable rotation of the Earth represent the nutation–precession in longitude and obliquity, respectively. The precession here is the so-called lunisolar precession reflecting the actual motions of the equator (or equivalently, of the figure axis) and differs from the general precession, which is defined relative to the moving ecliptic and therefore includes an ecliptic motion term (see the last paragraph of **Section 3.10.2.1.1**).

The torque equations in the principal axis frame are Euler equations for the angular velocity components. The angular velocity components Ω_x, Ω_y, Ω_z in the principal axis frame, which Bretagnon et al. (1997, 1998) denote by (p, q, r), are related to $\dot{\psi}$, $\dot{\omega}$, $\dot{\varphi}$ through the (exact) kinematic relations

$$\begin{aligned}\Omega_x &= \dot{\psi}\sin\omega\sin\tilde{\varphi} + \dot{\omega}\cos\tilde{\varphi}\\ \Omega_y &= \dot{\psi}\sin\omega\cos\tilde{\varphi} - \dot{\omega}\sin\tilde{\varphi}\\ \Omega_z &= \dot{\psi}\cos\omega + \dot{\varphi}\end{aligned} \qquad [116]$$

On introducing these expressions for the angular velocity components into the Euler equation [11], one obtains a set of coupled second-order differential equations for the Euler angles, which involve small nonlinear terms arising from the small deviations of ω and $\dot{\varphi}$ from their constant mean values. The torque components that appear on the right-hand side of the equations are in the principal axis frame. Their expressions may be obtained from the torque components referred to the ITRF, which are given in eqns [59] and [64]. For the complex combination of the equatorial components, all one needs to do is to multiply the former expression by $e^{i\alpha}$, where α is the longitude of the x-axis of the principal axis frame in the ITRF (see eqn [63]). The axial component of the torque is the same in both frames. Of course, the expressions in the preceding text have to be summed over n and m to get the torque due to the full potential of the external body.

The time dependence of the torque produced by the Moon, Sun, or planet is obtainable by using the analytic expressions given in the relevant ephemerides for the position coordinates of that body. The coupled differential equations may then be solved to any desired accuracy (given the Earth's structure parameters to sufficient accuracy) by an iterative process, which is needed because of the nonlinearities. Corrections for relativistic effects (Brumberg et al., 1992; Fukushima, 1991; Klioner, 1998; Soffel

and Klioner, 1998) have to be applied to the solutions thus obtained. The final nutation series constructed in this manner by Bretagnon et al. (1997, 1998) is named SMART97.

3.10.9 Axially Symmetrical Ellipsoidal Nonrigid Earth: Torque Equations and Solutions

In the foregoing sections, we considered the wobble and nutation in some simplified models, then established a number of basic results of general applicability, and briefly dealt with treatments of the nutations of the rigid Earth. We turn now to a study of more realistic models having one of more of the following features: deformability, existence of a fluid core and SIC, etc. The torque approach will be employed for these studies.

We observe at the outset that the basic equation of motion governing the rotation variations as seen from the surface of the Earth in relation to a terrestrial frame continues to be the torque equation [8] for the whole Earth; the angular velocity $\boldsymbol{\Omega}$ appearing in the equation is that of the mantle. An additional equation is needed to take account of the rotation variations of the fluid core, which are coupled to, but not the same as, those of the mantle. Two more equations are needed when an SIC is present, as will be seen shortly.

3.10.9.1 Deformable Wholly Solid Earth

If the Earth were wholly solid, the only difference from the rigid case would be in the inertia tensor. The deformation from the direct action of the tidal potential, as well as indirect effects like the perturbations of the centrifugal resulting from Earth rotation variations, has the consequence that the inertia tensor [C] is no longer diagonal. So the angular momentum components become

$$\begin{pmatrix} H_x \\ H_y \\ H_z \end{pmatrix} = \begin{pmatrix} A + c_{11} & c_{12} & c_{13} \\ c_{21} & A + c_{22} & c_{+} \\ c_{31} & c_{32} & C + c_{33} \end{pmatrix} \begin{pmatrix} m_x \Omega_0 \\ m_y \Omega_0 \\ (1 + m_z)\Omega_0 \end{pmatrix} = \Omega_0 \begin{pmatrix} Am_x + c_{13} \\ Am_y + c_{23} \\ C(1 + m_z) + c_{33} \end{pmatrix} \quad [117]$$

with the neglect of second-order terms, which involve the product of any of the C_{ij} with m_x or m_y or m_z. The components of $\boldsymbol{\Omega} \times \mathbf{H}$ may be obtained using the previously mentioned. Its components, leaving out the second-order terms, are $(Cm_y - (Am_y + c_{23}),\ -Cm_x + (Am_x + c_{13}),\ 0)$. When these are substituted into the torque equation $\mathrm{d}\mathbf{H}/\mathrm{d}t + \boldsymbol{\Omega} \times \mathbf{H} = \boldsymbol{\Gamma}$, one obtains the following equations:

$$\begin{aligned} \Omega_0(A\dot{m}_x + \Omega_0 \dot{c}_{13}) + (C - A)\Omega_0^2 m_y - \Omega_0^2 c_{23} &= \Gamma_x \\ \Omega_0(A\dot{m}_y + \Omega_0 \dot{c}_{23}) - (C - A)\Omega_0^2 m_y - \Omega_0^2 c_{13} &= \Gamma_y \\ C\Omega_0 \dot{m}_z + \dot{c}_{33} &= \Gamma_z \end{aligned} \quad [118]$$

Now, we take the complex combination of the first two equations and denote $c_{13} + ic_{23}$ by $\tilde{c}_3$. Then, on using for the torque the expression [50] with $e' = 0$ (for axial symmetry), we obtain the following equation:

$$A\frac{\mathrm{d}\tilde{m}}{\mathrm{d}t} - ie\Omega_0 A\tilde{m} + \frac{\mathrm{d}\tilde{c}_3}{\mathrm{d}t} + i\Omega_0 \tilde{c}_3 = -ieA\Omega_0 \tilde{\phi} \quad [119]$$

For a spectral component with frequency $\sigma\Omega_0$, we have then

$$A(\sigma - e)\tilde{m}(\sigma) + (\sigma + 1)\tilde{c}_3(\sigma) = -eA\tilde{\phi}(\sigma) \quad [120]$$

In order to solve this equation for $\tilde{m}(\sigma)$, we need to know how $\tilde{c}_3$ is related to the wobble $\tilde{m}$ and to the external forcing $\tilde{\phi}$. One can show from deformation theory that the contribution to $\tilde{c}_3$ is $-A\kappa\tilde{\phi}$. Since the value of κ could be frequency-dependent, as we shall see later, we take this relation to be in the frequency domain. The parameter κ is a measure of the deformability of the Earth under forcing by $\tilde{\phi}$ and is referred to as a compliance; it is closely related to the Love number k of the unperturbed Earth:

$$\kappa = \frac{\Omega_0^2 a^5}{3GA} k = \frac{ek}{k_\mathrm{f}}, \quad k_\mathrm{f} = \frac{3GAe}{\Omega_0^2 a^5} \quad (121)$$

where A is the mean radius of the Earth and G is the constant of universal gravitation, $G = 6.67259 \times 10^{-11}\ \mathrm{m^3\ kg^{-1}\ s^{-2}}$. (The Love number k is defined by the statement that the incremental gravitational produced at the surface of the Earth by the redistribution of matter within the Earth as a result of the action of the lunisolar potential is k times the lunisolar potential at the surface.) The so-called fluid Love number k_f plays the same role for a hypothetical wholly fluid Earth. Other Love numbers h and l relate the tide-induced vertical and horizontal displacements at the Earth's surface to the lunisolar potential. (In these cases, the displacement at the surface is h/g times the lunisolar potential.) There are also load Love numbers k', h', l', which pertain to the potential perturbation and the displacements caused by loading on the Earth's surface. The contribution from the incremental centrifugal potential (associated with the wobble $\tilde{m}$) to $\tilde{c}_3$ is $Ak\tilde{m}$, for the simple reason that the difference between the centrifugal potentials associated with the perturbed and unperturbed angular velocities $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ is identical to the tesseral potential given by eqn [31], except for the replacement of $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$ by $-m_x$, $-m_y$. Thus,

$$\tilde{c}_3 = Ak\left(\tilde{m} - \tilde{\phi}\right) \quad [122]$$

On substituting this expression for $\tilde{c}_3$ in eqn [120], the equation becomes

$$[\sigma - e + (\sigma + 1)k]\tilde{m}(\sigma) = -[e - (\sigma + 1)k]\tilde{\phi}(\sigma). \quad (123)$$

One sees, by setting $\tilde{\phi}(\sigma) = 0$, that the frequency $\sigma_\mathrm{E} = e$ of the Euler free wobble mode of the rigid Earth is now replaced by $\sigma_1 = (e - k)/(1 + k)$. It turns out that k is close to $e/3 \approx 0.001$, and so $\sigma_1 \approx (2/3)\sigma_\mathrm{E}$, meaning that the period of the free wobble of the deformable Earth with no fluid regions is approximately 50% longer than that of the rigid Earth. As for the wobble response to forcing by $\tilde{\phi}(\sigma)$, the change from the rigid case consists in a replacement of e by the frequency-dependent effective value $e - (\sigma + 1)k$. It becomes clear, on recalling that $(1 + \sigma)$ lies within the low-frequency band, that the modification becomes larger as a moves away from -1 but is never very large, as $|\sigma + 1|$ is at most of the order of 0.1 for spectral components of any significance.

3.10.9.2 Two-Layer Earth, with Mantle and Fluid Core

3.10.9.2.1 Equations of motion

When the rotation of the mantle is not about its symmetry axis, the fluid flow within the core consists of a rotation of the fluid as a whole about yet another axis, plus a residual flow needed to make the flow at the ellipsoidal CMB normal to the boundary. The angular velocity of the rotational flow is also not along the symmetry axis; we denote it by $\boldsymbol{\Omega}_f$. The differential angular velocity between the core and the mantle is then

$$\omega_f = \boldsymbol{\Omega}_f - \boldsymbol{\Omega} \qquad [124]$$

with components $\Omega_0(m_{fx}, m_{fy}, m_{fz})$. (Superscripts or subscripts f are used to identify quantities pertaining to the fluid core. m_{fx} and m_{fy} represent the differential wobble between the core and the mantle.) The contribution $[\mathbf{C}]_f \cdot \omega_f$ to the angular momentum of the Earth from this differential rotation has to be added now to $[\mathbf{C}] \cdot \boldsymbol{\Omega}$ to get the total $\mathbf{H}$. Consequently, the $[\mathbf{C}]_f \cdot (d\omega_f/dt)$ and $\boldsymbol{\Omega} \times ([\mathbf{C}]_f \cdot \omega_f)$ will appear on the left-hand side of the torque equation in addition to the terms in $[\mathbf{C}]$ that we had earlier. The components of these can be readily evaluated, noting that the components of $[\mathbf{C}]_f \cdot \omega_f$ are $(A_f m_{fx}, A_f m_{fy}, C_f m_{fz})$ to the first order in small quantities. On using these, we find the new equation of motion to be

$$A\Omega_0 \frac{d\tilde{m}}{dt} - ieA\Omega_0^2 \tilde{m} + A_f \Omega_0 \frac{d\tilde{m}_f}{dt} + iA_f \Omega_0^2 \tilde{m}_f + \Omega_0 \frac{d\tilde{c}_3}{dt} + i\Omega_0^2 \tilde{c}_3 = \tilde{\Gamma} \qquad [125]$$

This is no longer an equation for $\tilde{m}$ alone: it involves $\tilde{m}_f$ also.

A second equation for $\tilde{m}$ and $\tilde{m}_f$ is provided by the torque equation for the fluid core alone. Unlike the equation for the whole Earth, wherein the internal interactions between its different parts were of no relevance, the equation for the core alone must include torques exerted by all bodies outside itself, namely, the mantle as well as the celestial bodies. The obvious interaction mechanism comes from the impact of the flowing fluid core on the CMB (which is nonspherical) and the resulting inertial reaction from the mantle on the core, when the axis of the global rotation of the core is not aligned with the symmetry axis. This reaction is reflected in the perturbations of the various dynamical quantities within the fluid, like the pressure, gravitational potential, and density. The combined effect of all this is an inertial torque on the core that is at the first order proportional to the differential wobble ω_f and the dynamical ellipticity e_f of the core. (e_f is defined, analogously to e, as $(C_f - A_f)/A_f$.) Mathematically, this result comes from a proof that the total torque $\boldsymbol{\Gamma}_f$ on the fluid core can be reduced to $(\boldsymbol{\Omega} + \omega_f) \times H_f$ to the first order in the surface flattening of the CMB. The torque arises from forces derived from the pressure P and the gravitational potential (of the Earth itself as well as of celestial bodies) and is represented by the integral

$$\boldsymbol{\Gamma}_f = -\int r \times \left(\nabla P + \rho \nabla \phi_g\right) dV \qquad [126]$$

where the integration is over the instantaneous volume of the core. The proof is not very simple and will not be presented here (see, for instance, the annex in Mathews et al., 1991). Given this result, the torque equation for the core, which is of exactly the same form as eqn [8] except for the replacement of $\mathbf{H}$ by $\mathbf{H}_f$ and $\boldsymbol{\Gamma}$ by $\boldsymbol{\Gamma}_f$, reduces to

$$\frac{d\mathbf{H}_f}{dt} - \omega_f \times \mathbf{H}_f = 0 \qquad [127]$$

This is the equation in reduced form for the angular momentum of the core. An equivalent form was derived by Poincaré (1910) by taking the fluid core to be homogeneous and incompressible assuming for the flow velocity in the fluid a form that is linear in x, y, z and satisfies, from the outset, the condition of being normal to the core boundary. A derivation without making such simplifying assumptions was given in elegant form by Sasao et al. (1980), using ideas based on a much more complicated derivation by Molodensky (1961).

To obtain the equation of motion for $\tilde{m}_f$, we need to express the preceding equation in component form. The components of $\mathbf{H}_f$ may be obtained in the same manner as those of $\mathbf{H}$:

$$\mathbf{H}_f = \Omega_0 \begin{pmatrix} A_f(m_x + m_{fx}) + c_{13}^f \\ A_f(m_y + m_{fy}) + c_{23}^f \\ C_f(1 + m_z + m_{fz}) + c_{33}^f \end{pmatrix} \qquad [128]$$

Using these to express eqn [127] in terms of components, and then taking the complex combination of the first two components, we obtain

$$A_f \Omega_0 \dot{\tilde{m}} + A_f \Omega_0 \dot{\tilde{m}}_f + i(1 + e_f) A_f \Omega_0^2 \tilde{m}_f + \Omega_0 \dot{\tilde{c}}_{3f} = 0 \qquad [129]$$

The pair of coupled eqns [125] and [129] govern the temporal variations of $\tilde{m}$ and $\tilde{m}_f$. On going over to the frequency domain, the dynamical variables to be solved for are $\tilde{m}(\sigma)$ and $\tilde{m}_f(\sigma)$; and c_3 and c_{3f} go over into $c_3(\sigma)$ and $c_{3f}(\sigma)$. The last two quantities have to be written, as before, in terms of the forcing and the dynamical variables. We now have, in the frequency domain,

$$\tilde{c}_3 = A\left[k\left(\tilde{m} - \tilde{\phi}\right) + \xi \tilde{m}_1\right]$$
$$\tilde{c}_{3f} = A_f\left[\gamma\left(\tilde{m} - \tilde{\phi}\right) + \beta \tilde{m}_f\right] \qquad [130]$$

The terms proportional to $\tilde{m}_f$ represent the deformations due to the centrifugal effect of the differential wobble of the core. There are now four compliances in all, representing the deformabilities of the Earth and of the core under the centrifugal forcing due to wobbles of the whole Earth or due to the differential wobbles of the core. All of them are of order 10^{-3} or smaller, though the precise values depend on the Earth model used:

$$k \approx 1.0 \times 10^{-3}, \quad \xi \approx 2 \times 10^{-4}$$
$$\gamma \approx 2 \times 10^{-3}, \quad \beta \approx 6 \times 10^{-4} \qquad [131]$$

On substituting the previous expressions into the spectral components of eqns [125] and [129], we obtain the following coupled equations:

$$[\sigma - e + (\sigma + 1)k]\tilde{m}(\sigma) + (\sigma + 1)(A_f/A + \xi)\tilde{m}_f(\sigma) = -[e - (\sigma + 1)k]\tilde{\phi}(\sigma) \qquad [132]$$

$$\sigma(1 + \gamma)\tilde{m}(\sigma) + (\sigma + 1 + \beta\sigma + e_f)\tilde{m}_f(\sigma) = \sigma\gamma\tilde{\phi}(\sigma) \qquad [133]$$

An important point to note is that for $\sigma = -1$, which corresponds to zero frequency relative to the CRF, the first of the preceding equations reduces to $(\sigma - e)\tilde{m}(\sigma) = -e\tilde{\phi}(\sigma)$, which is precisely the same as for the rigid Earth. This is an example of

a general property, which Poincaré named *gyrostatic rigidity*: at zero frequency in an inertial frame, the nutation of a nonrigid Earth and the associated wobble in the terrestrial frame are the same as for a rigid Earth with the same value of e.

3.10.9.2.2 Wobble normal modes

We are now in a position to determine the wobble normal modes and eigenfrequencies of the two-layer Earth by solving the homogeneous equations obtained by setting $\tilde{\phi}(\sigma) = 0$. To the lowest order in the ellipticities and compliance parameter, we find the eigenfrequencies to be

$$\sigma_1 = \sigma_{\mathrm{CW}} = \frac{A}{A_{\mathrm{m}}}(e-k)$$
$$\sigma_2 = \sigma_{\mathrm{NDFW}} = -\left(1 + \frac{A}{A_{\mathrm{m}}}(e_{\mathrm{f}} - \beta)\right) \qquad [134]$$

where A_{m} is the moment of inertia of the mantle: $A_{\mathrm{m}} = A - A_{\mathrm{f}}$. The labels CW and NDFW stand for 'CW' and 'NDFW,' respectively. The CW frequency of the two-layer Earth differs from that of the wholly solid Earth just by a factor (A/A_{m}), which is about 9/8. As for the NDFW, its frequency is $-1-1/430$ cpsd, justifying the characterization as nearly diurnal. (In nutation, this mode appears as the FCN, or more precisely, the retrograde free core nutation (RFCN), with the retrograde frequency $-(A/A_{\mathrm{m}})(e_{\mathrm{f}}-\beta) \approx -1/430$ cpsd.) This mode has a very important role to play in the forced wobbles, because of its location in the middle of the retrograde diurnal band in which the forcing frequencies lie: Its presence results in resonant enhancement of several of the prominent wobbles and the corresponding nutations. In particular, the amplitude of the retrograde annual nutation with frequency $\approx -1/366$ cpsd is enhanced by over 30% relative to that of the same nutation of the rigid Earth, because its frequency σ in the terrestrial frame is $\approx -1-1/366$ cpsd, very close to the FCN eigenfrequency.

3.10.9.2.3 Solution of the wobble equations: resonances and nutation amplitudes

Solution of the pair of eqns [132] and [133] with the forcing potential $\tilde{\phi}(\sigma)$ present yields the forced wobble $\tilde{m}(\sigma)$ of the mantle and the differential wobble $\tilde{m}_{\mathrm{f}}(\sigma)$ of the fluid core. To the first order in the compliances and ellipticities, one finds that

$$\tilde{m}(\sigma)\frac{(\sigma+1)[eA + \gamma\sigma A_{\mathrm{f}} - (1+\sigma)kA]}{A_{\mathrm{m}}(\sigma-\sigma_1)(\sigma-\sigma_2)}\tilde{\phi}(\sigma) \qquad [135]$$

The two factors appearing in the denominator give rise, not surprisingly, to resonances associated with the eigenfrequencies σ_1 and σ_2. The expression for the amplitude $\tilde{m}_{\mathrm{f}}(\sigma)$ of the differential wobble of the fluid core may be similarly written down.

The amplitude of the nutation associated with the wobble [135] may be obtained by applying the kinematic relation $\tilde{\eta}(\nu) = -\tilde{m}(\sigma)/\nu$, with $\nu = 1+\sigma$. Given the values of the Earth parameters, numerical evaluation may be done using $\tilde{\phi}(\sigma)$ obtained from a tide table.

Alternatively, one may take the transfer function $T_{\mathrm{w}}(\sigma) = T_{\mathrm{n}}(\nu)$ obtained by dividing the preceding expression by the rigid Earth wobble amplitude and $\tilde{m}_{\mathrm{R}}(\sigma) = e\tilde{\phi}(\sigma)/(e-\sigma)$ by multiplying it by rigid Earth nutation amplitude for the frequency $\nu = \sigma + 1$, for example, from REN2000 of Souchay et al. (1999), to obtain the nutation amplitude for the nonrigid Earth. In this case, one does not need to use the amplitude of the tidal potential. This is the approach adopted by Mathews et al. (2002).

3.10.9.3 Coupling of the Core and the Mantle at the CMB

It was mentioned in the beginning of Section 3.10.9.2.1 that the pressure and gravitational forces acting on the fluid core result in an effective inertial coupling between the core and the mantle, which is proportional to the dynamical ellipticity of the core. No other couplings were included in the equations of that section. The efforts to refine nutation theory in order to achieve close agreement between theoretical predictions and the high-precision observational data available now have led to the investigation of other coupling mechanisms that might be significant, especially EM and viscous couplings at the CMB.

The magnetic fields generated in the fluid core by the geodynamo mechanism extend into the mantle across the CMB. As differential motions between the solid and fluid sides of the CMB take place as a result of the differential wobble between the mantle and the fluid core, the field lines remain 'frozen' in the fluid because of the very high conductivity of the fluid and move with it, and the continuing portion of the lines on the mantle side tends to sweep through the mantle material. But if there is a highly conducting layer at the bottom of the mantle, the attempted motion of the field lines induces currents, which, on the one hand, interact with the magnetic field and produce a Lorentz force on the matter and produce a magnetic field that increments the main magnetic field $\mathbf{B}_0$, which existed in the absence of the wobbles. The Lorentz forces all through the conducting layer result in a torque on the mantle and an equal reaction by the mantle on the fluid core. The strength of this torque that couples the two regions depends on the conductivity of the mantle layer, on the frequency of the wobble causing the relative motion, and on the strength and configuration of the main field over the surface of the CMB. If the field is sufficiently weak, the torque is 45° out of phase with the wobble and has a strength proportional to the mean squared of the radial component B_{r} of $\mathbf{B}_0$ over the boundary surface; otherwise, the proportionality is only approximate, and the phase difference too depends on the overall strength and configuration of the field. In any case, the equatorial torque on the fluid core can be expressed as

$$\tilde{\Gamma}^{\mathrm{CMB}} = -i\Omega_0^2 A_{\mathrm{f}} K^{\mathrm{CMB}} \tilde{m}_{\mathrm{f}} \qquad [136]$$

where K^{CMB} is a complex coupling constant depending on the physical quantities mentioned earlier. The viscosity of the fluid core is widely believed to be negligible; in case it is not, both EM and viscous effects can be encompassed under an expression of the same form as in the preceding text, with K^{CMB} becoming dependent also on the coefficient of viscosity. The mathematical derivation of the expression for K^{CMB} in terms of the relevant physical parameters will not be presented here. The interested reader may refer to Buffett (1996a, 1996b, 2010), Buffett et al. (2002), Mathews and Guo (2005), Koot et al. (2008, 2010), and Koot and Dumberry (2011).

A simple reflection will show that the effect of these couplings can be incorporated into the theory of Section 3.10.9.2 simply by adding K^{CMB} to the coefficient $\tilde{m}_{\mathrm{f}}$ in eqn [133].

This replacement does not affect σ_{CW} of the CW. But σ_{NDFW} does get modified; the change consists in the replacement of $(e_f - \beta)$ in eqn [134] by $(e_f - \beta + K^{CMB})$, making it complex. Consequently, the NDFW resonance acquires a width.

3.10.9.4 Anelasticity and Ocean Tide Effects

It has been known, from seismological data on the oscillational normal modes of the Earth and on the propagation of seismic waves, that the response of the material of the Earth's mantle to stresses is not strictly elastic. A small part of the response is not instantaneous; it decays over a finite though short period of time. This part of the response is said to be 'anelastic.' When transformed to the frequency domain, the anelastic response manifests itself through a frequency dependence of the rheological parameters like the Young's modulus and rigidity modulus; in addition, these parameters become complex. Wahr and Bergen (1986) showed that anelasticity effects are in the range of tenths of mas on the amplitudes (both in phase and out of phase) of a number of spectral components of nutation. The anelasticity model that they employed would produce increments $\Delta^{AE}k$, $\Delta^{AE}\gamma$, etc., proportional to

$$\left\{1 - \left(\frac{\omega_m}{\omega}\right)^{\alpha}\right\} \cot\frac{\alpha\pi}{2} + i\left(\frac{\omega_m}{\omega}\right)^{\alpha} \qquad [137]$$

to the various compliances. Here, ω is the frequency of the tidal potential responsible for the nutation of interest and ω_m is a reference frequency in the band of frequencies of the seismic waves used for estimating the elastic rheological parameters; it is taken to be equivalent to a period of 1, 200, or 300 s in different models (Dziewonski and Anderson, 1981; Widmer et al., 1991). The power law index α is typically considered to be around 0.15 (between 0.1 and 0.2). Evaluation of the proportionality constant has to be done by integration of the deformation equations for the Earth model.

The computation of the effects of loading of the Earth's crust by ocean tides, as well as those of ocean tidal currents, was first done by Wahr and Sasao (1981) using the theoretical formalism of Sasao and Wahr (1981). They included the effect of ocean tides in eqns [125] and [129] by adding the ocean tidal contributions $\tilde{c}_3^o$ and $\tilde{c}_{3f}^o$ to $\tilde{c}_3$ and $\tilde{c}_{3f}$, respectively, where the added quantities reflect the changes in the inertia tensors of the whole Earth and of the fluid core due to the deformation produced by the tidal redistribution of ocean mass, together with the equivalent of the angular momentum content of the tidal currents in the oceans. Using estimates of these quantities obtained with the help of ocean tidal models, they found increments of up to about 1 mas to the amplitudes of the leading nutations. Mathews et al. (2002) expressed $\tilde{c}_3^o$ and $\tilde{c}_{3f}^o$ in terms of incremental compliances $\Delta^{OT}k$ and $\Delta^{OT}\gamma$. These were computed using empirical formulas (separately for the mass redistribution part and for the tidal current part) representing their frequency dependences; the parameters in the formula were estimated with the help of available data, in particular, the ocean tide angular momentum data provided by Chao et al. (1996) for four prominent tidal components; the reference in the preceding text may be consulted for details. The results that Mathews et al. (2002) obtained for the increments to the nutations from ocean tidal (starting from the results of Chao et al., 1996) as well as anelasticity effects are close to those of earlier authors, for example, Wahr and Sasao (1981) and Wahr and Bergen (1986).

3.10.9.5 Inclusion of SIC

In the presence of the SIC, the fluid layer is more appropriately referred to as the FOC; and all the quantities labeled by the subscript f now refer to the FOC. Quantities pertaining to the SIC will be identified by the subscript s. The SIC has a moment of inertia A_s that is only 1/1400 of that of the whole Earth, and so it would seem that it could not influence the rotation of the Earth as a whole to any significant extent. (Note that the mantle moment of inertia is $A_m = A - A_f - A_s$ now.) Recent investigations, beginning with Mathews et al. (1991), De Vries and Wahr (1991), Dehant et al. (1993), and Legros et al. (1993), have revealed however that a new NDFW appears as a result of the role of the inner core in the dynamics and that the associated resonance has an appreciable effect on the amplitudes of a few nutations.

The inner core is floating within the fluid core, and its symmetry axis is free to go out of alignment with that of the mantle. Consequently, two new equations enter the picture. One is the torque equation governing the time variations of the differential wobble $\tilde{m}_s$ of the SIC, that is, of the complex combination of the equatorial components of $(\boldsymbol{\Omega}_s - \Omega)/\Omega_0$; the other is a kinematic equation relating $\tilde{m}_s$ to $\tilde{n}_s$, the complex version of equatorial components of the offset $(\mathbf{i}_s - \Omega_0/\Omega_0)$ of the inner core symmetry axis $\mathbf{i}_s$ from that of the mantle. Thus, one has a set of four equations, which couple the variables $\tilde{m}$, $\tilde{m}_f$, $\tilde{m}_s$, $\tilde{n}_s$ to one another.

The torque equation involves gravitational and other torques acting on the inner core. The gravitational torque is made up of the torque exerted by the celestial body and a torque arising from the misalignment between the ellipsoidal inner core and the rest of the Earth (which is also of ellipsoidal symmetry except for a slight realignment of the inner core boundary (ICB) because of the tilt of the inner core). Both parts are proportional to the ellipticity e_s of the SIC; the internal gravitational part is proportional also to the density contrast between the SIC and the fluid core just outside the ICB and involves a gravitational coupling constant α_g between the SIC and the rest of the Earth. An additional torque arises from EM (and possibly viscous) couplings at the ICB. It may be expressed as $\Gamma^{ICB} = -i\Omega_0^2 A_s K^{ICB} \tilde{m}_s$; the complex ICB coupling constant K^{ICB} appears in the torque equations for both regions coupled by it, namely, the FOC and the SIC.

The expression for $\mathbf{H}$ now gets an additional contribution from $\tilde{m}_s$, and the moments of inertia of all the regions bear the imprint of the misalignment $\tilde{n}_s$. Furthermore, $\tilde{c}_3$, $\tilde{c}_{3f}$, and the new $\tilde{c}_{3s}$ all involve terms proportional to $\tilde{m}_s$ besides those in $\tilde{m}$ and $\tilde{m}_f$, and a number of new compliances enter the picture.

3.10.9.5.1 Solution of the equations for the three-layer Earth

Solution of the four coupled linear equations with the tidal potential $\tilde{\phi}$ set equal to zero yields two new normal modes in addition to the earlier two. One has a frequency depending on e_s and the ratio of the density contrast mentioned earlier to the mean density of the inner core; it is very close to −1 cpsd.

The corresponding nutation, with a very small prograde frequency, is referred to as the PFCN or FICN. The resonance associated with this mode has a nontrivial effect on forced nutations at nearby frequencies. The frequency of the other wobble normal mode, called the inner core wobble (ICW) mode, is almost an order of magnitude smaller than that of the CW, and its effect is quite ignorable.

The forced wobbles $\tilde{m}$, $\tilde{m}_f$, $\tilde{m}_s$ as well as the offset $\tilde{n}_s$ of the inner core axis for any excitation frequency σ are found by solving the set of four equations with $\tilde{\phi}(\sigma)$ taken to be nonzero. Division of $\tilde{m}(\sigma)$ by the rigid Earth amplitude $\tilde{m}_R(\sigma)$ from eqn [55] gives the transfer function $T_w(\sigma) = T_n(\nu)$, and multiplication of this quantity by the rigid Earth amplitude $\tilde{\eta}_R(\nu)$ of the nutation of frequency $\nu = \sigma + 1$ finally gives the nutation amplitude $\tilde{\eta}(\nu)$ for the nonrigid Earth.

The magnitude to which the resonance associated with each normal mode affects the amplitudes of forced nutation for a given Earth model may be displayed through the resonance expansion of the transfer function for that model. For the three-layered Earth, which has four eigenmodes, the expansion takes the form

$$T_w(\sigma) = R + R'(1+\sigma) + \frac{R_{CW}}{\sigma - \sigma_{CW}} + \frac{R_{FCN}}{\sigma - \sigma_{FCN}} + \frac{R_{FICN}}{\sigma - \sigma_{FICN}} + \frac{R_{ICW}}{\sigma - \sigma_{ICW}} \quad [138]$$

The form of the expansion in the preceding text presumes that the dynamical flattening used in the construction of the rigid Earth nutation series accurately represents the flattening of the real Earth. It was found by Mathews et al. (2002), on requiring the best fit of the results of nonrigid Earth nutation theory to observational data, that a slightly different value of the flattening parameter is called for. It is necessary then to modify the transfer function accordingly (see Mathews et al., 2002, for more detail).

3.10.9.6 Confronting Theory with Observations

The main aim of the nutation theory is to provide theoretical models from which the nutation amplitudes/coefficients can be computed accurately, making it possible to match the observationally estimated series by the computed nutation series at about the same level as the uncertainties in the observational estimates (a couple of tenths of mas) when observational inputs are used for the inherently unpredictable contributions to variations in orientation related to excitations of geophysical origin (atmospheric and nontidal ocean effects, the free wobbles and nutations, etc.). Such a theoretical model, combined with an accurate model for LOD variations, can then be used to make accurate predictions for the Earth's orientation in space at future epochs.

For highly accurate computations to be possible, one needs sufficiently accurate values of the relevant Earth parameters. The other essential input, namely, accurate values of the amplitudes of the components of the tidal potential or, alternatively, the values of the rigid Earth nutation amplitudes, is provided by existing theories. Given all these, the precession and low-frequency nutation due to the action of the degree 2 tesseral gravitational potential of the solar system bodies at the first order in the potential may be calculated to high accuracy using the theory outlined in Section 3.10.9 as applied to a three-layer Earth, taking into account the effects dealt with in Section 3.10.9.3 and subsequent sections. However, to get a complete accounting of the nutations, one needs to add some other contributions:

(i) Contributions to nutations and precession from torques of the second order in the perturbing potential, which arise from the action of each of the three parts (zonal, tesseral, and sectorial) of the potential on the increment to the density function $\rho(\mathbf{r})$ that is produced by the tidal deformation produced by the other two parts. These have been evaluated in part by Folgueira et al. (1998a, 1998b), Souchay and Folgueira (2000), and Mathews et al. (2002). A complete calculation done since by Lambert and Mathews (2005) shows that mutual cancelations limit the total effect to about 30 μas on the 18.6-year nutation (about a third of the earlier result) and that the effect on precession is practically nil.

(ii) Geodesic precession and nutation, arising from the difference between kinematic and dynamical quantities in general relativity. This relativistic contribution is about -19.2 mas year^{-1} to the precession and about -30 and $+30$ μas, respectively, to the prograde and retrograde annual nutations; the effect on other nutations is much smaller.

(iii) The contribution from atmospheric pressure tides generated by solar thermal effects that is of the order of 0.1 mas at the prograde annual period and possibly nonnegligible at a couple of other periods.

(iv) Contributions of the order of 0.1 mas or smaller to a couple of low-frequency nutations from the degree 3 tidal potential acting on J_3 and about 25 μas year^{-1} to precession from the degree 4 potential acting on J_4 (see, for instance, Souchay et al., 1999; Williams, 1994).

Other variations in Earth rotation with spectra outside the low-frequency band in the terrestrial frame are produced by potentials of degrees >2 and of orders $m \neq 1$; they are grouped under polar motions of the CIP and are not considered as nutations. For a treatment of these, see Mathews and Bretagnon (2003).

Contributions from the effects (i)–(iv) in the preceding text form part of the observed nutation and precession. These contributions have to be subtracted out from the observational estimates of the precession rate and the coefficients/amplitudes of low-frequency nutations, before comparing to the results of computations based on the theory of Section 3.10.9.

The values of the Earth parameters appearing in the theory are obtained by computations based on an Earth model. Only very rough estimates can be made of the EM coupling constants K^{CMB} and K^{ICB}, assuming some model for the structure and strengths of the core magnetic fields crossing the CMB and ICB. If, on using these parameter values, computational results on the precession rate and nutation amplitudes turn out to have unacceptably large residuals relative to their observational estimates, one is forced to conclude that the value(s) of one or more of the Earth parameters do not reflect the properties of the real Earth accurately enough, implying that the detailed Earth model used for computation of the parameter values needs to be suitably modified or refined. One has then the

option of adjusting the values of those parameters to which the offending residuals are most sensitive. Such a course was adopted by Gwinn et al. (1986) and Herring et al. (1986), and it led to the inference that the ellipticity e_f of the core has to be about 5% higher than its Earth model value. This deduction was based on the observation that the retrograde annual nutation, in which a large residual of about 2 mas was found, is strongly influenced by the FCN resonance as the FCN eigenfrequency is quite close to the retrograde annual and that a rather small adjustment of e_f would change the eigenfrequency enough to eliminate the residual. The adjustment needed to e_f has since been modeled as a result of mantle convection, which causes the flattening of the CMB to deviate from the hydrostatic equilibrium value (Defraigne et al., 1996), an effect that is not taken into account in Earth models constructed from radial seismic data taken together with the assumption of hydrostatic equilibrium. It had been noted even earlier by Wahr (1981), from observational data on the precession rate, that the ellipticity e of the whole Earth has to be higher than the hydrostatic equilibrium value by over 1%.

In recent versions of torque equations, which take account of the inner core as well as the couplings at the CMB and the ICB between the fluid core and the adjoining solid regions, there are parameters like K^{CMB} and K^{ICB} for which no inputs are available from Earth models though the estimates of the mean squared magnetic field at the CMB inferred from satellite measurements of magnetic fields outside the Earth make an initial estimate for K^{CMB} possible. Confronting nutation theory with observational data provides a means of obtaining estimates with relatively small uncertainties for such parameters.

In constructing the nutation model MHB2000 on which the current IAU model IAU2000A is based, Mathews et al. (2002) employed a least squares procedure to obtain the best fit of the results of their theory to estimates of nutation and precession (see Herring et al., 2002) from a long VLBI data set, which was then up-to-date. Before the fitting, the small contributions listed under items (i)–(iv) in the preceding text were subtracted from the observational estimates, for reasons explained earlier. The least squares process involves repeated adjustments and optimization of the values of a few Earth parameters, namely, the ellipticities e, e_f, compliances κ, γ, the imaginary part of K^{CMB}, and the real and imaginary parts of K^{ICB}, while retaining values based on the PREM Earth model for all the other parameters. (The real part of K^{CMB} appears in combination with e_f in the equations of the theory, and the two could not be independently estimated. Instead, a theoretical relation between the real and imaginary part of K^{CMB} has been used, and e_f only has been estimated.) The theoretical values for the fit were obtained from the solution of the equations for the three-layer Earth as outlined in **Section 3.10.9.5.1**, including the various effects considered in the sections preceding it. The rigid Earth nutation amplitudes used to multiply the transfer function were taken by these authors from the REN2000 series of Souchay et al. (1999). In carrying out the numerical evaluation of the nutation amplitudes and the precession rate, an a priori set of values was employed for the Earth parameters occurring in the equations. The values of the selected Earth parameters were optimized by a fitting process, which seeks a least squares minimization of the residuals between the observational estimates and the theoretically computed values. The process was iterative in view of the nonlinear dependence of the nutation amplitudes on the parameters being optimized. Within each step of the iterative process, another iterative loop was necessary. The reason was that the strong frequency dependence of the ocean tide increments $\Delta^{OT}\kappa$, $\Delta^{OT}\gamma$, etc., to the compliance parameters that appear in the set of wobble equations was not known a priori; it had to be determined through an ocean response model involving free parameters as well as frequency-dependent body tide and load Love numbers, which depend on the wobbles, which, in turn, have to be found by solving the wobble equations, and these equations involve the incremental compliances just mentioned, which we are trying to determine. The ocean responses known observationally for a very small number of tidal frequencies served as the input in this inner loop. In this process, the mutual influences of nutation–wobble, solid Earth deformation, and ocean tidal responses all come into full play in determining each type of response.

The final results are a set of optimized values for the Earth parameters chosen for adjustment by the process of arriving at the 'best fit' between the numbers from theory and observational data, on the one hand, and the MHB2000 nutation series on the other. The series is obtained from the nutation series that is output at the end of the iteration process (i.e., from the equations containing the optimized values for the adjusted parameters) by adding the small contributions, which had been subtracted out earlier.

The best-fit values of e and e_f are higher than their hydrostatic equilibrium values; these are interpreted as effects of mantle convection, as mentioned earlier. The excess (nonhydrostatic) part of e_f corresponds to an extra difference of a little under 400 m between the equatorial and polar radii of the CMB. The estimates obtained for K^{CMB} and K^{ICB} call for higher root-mean-squared (rms) values for the magnetic fields at the CMB and ICB than what are suggested by geodynamo models in general and also, in the case of the CMB, higher than the extrapolated values from satellite observations of magnetic fields along the satellite's orbit. No independent observations are available for the fields at the ICB.

If the observed complex coupling constant K^{CMB} is assumed to result from a purely EM coupling at the CMB, then it can be explained by an rms strength of the magnetic field at the CMB of the order of 0.7 mT, if the electrical conductivity of the outer core and of the lowermost mantle is fixed to 5×10^5 S m^{-1} (Buffett et al., 2002; Koot et al., 2010). For lower values of the conductivities, the rms strength of the field must be higher. From observations of the surface magnetic field, the large-scale components of the field (spherical harmonic degrees smaller than 14) at the CMB are known to have an rms strength of the order of 0.3 mT. Therefore, nutation observations suggest that a large part of the energy of the magnetic field at the CMB is contained in the small-scale components (harmonic degrees larger than 14), which are unobservable at the Earth's surface.

The possibility that the coupling constants might include the effect of viscous drag at the boundaries of the fluid core has been considered by Mathews and Guo (2005), Deleplace and Cardin (2006), Koot et al. (2010), and Koot and Dumberry (2011). These studies of viscomagnetic coupling at the CMB show that the viscous torque becomes significant only for values of the outer core viscosity of the order of 10^{-2} m^2 s^{-1}.

However, these values are in disagreement with inferences obtained from other sources (such as laboratory experiments on iron alloys at high pressure and temperature and theoretical 'ab initio' computations; see Alfè et al., 2000; Rutter et al., 2002), even if the viscosity is assumed to be an effective eddy viscosity due to turbulent motions (Buffett and Christensen, 2007). For realistic values of the outer core viscosity (i.e., lower than 10^{-4} m^2 s^{-1}), the viscous coupling does not contribute significantly to the observed dissipation.

Recently, another mechanism has been proposed to explain the observed coupling constant at the CMB. This mechanism is a variant of the EM coupling that requires the existence of a stratification of the light elements at the top of the core and of a topography at the CMB (Buffett, 2010). If the stratification at the top of the core is strong (with a buoyancy frequency $N = 0.09$ s^{-1}), this mechanism is able to explain the observed dissipation with an rms strength of the magnetic field at the CMB of 0.5 mT, with a lowermost mantle conductivity of 10^3 S m^{-1} and a CMB topography with a height of 120 m and horizontal length scale of 10^5 m.

In the nutation model adopted by the IAU in 2000, the only topographic feature that is considered in the nutation computation is the flattening of the CMB and the hydrostatic flattening of ICB. In reality, however, the CMB shape is bumped. Wu and Wahr (1997) had computed the effects of the topography on the nutation and had shown that some of the topography coefficients may induce large contributions (at a few tenths of mas level). This is due to the fact that some of the nutation frequencies may be close to inertial waves frequencies and are thus in resonance (amplified). However, these computations do not consider the presence of an inner core, which may alter the resonance frequencies.

3.10.10 Nutation–Precession from the Displacement Field

3.10.10.1 Displacement Field Approach

Perturbation by the tidal potential produces displacements of elements of matter from their equilibrium positions in the otherwise uniformly rotating Earth. The field of these displacements over the whole interior of the Earth may be analyzed into deformational and rotational parts; the first part is manifested as Earth tides and the second as rotation variations. The formulation and solution of the equations that govern the displacement field enable one to determine the precessional and nutational motions from the rotational part of the displacement field.

The approach of computing Earth orientation variations from the equations governing the field of displacements produced within the Earth by the tidal potential has been developed by Smith (1974), Wahr (1981), Dehant (1987a, 1987b), Dehant (1990), and Rogister (2001). This approach relies on detailed seismic models of the matter density and rheological parameters as functions of position within the Earth's interior; it can yield only accuracies commensurate with those of available Earth models.

3.10.10.1.1 Reference system used

The reference frame used here has its origin at the geocenter (center of mass of the Earth) and is tied to the hydrostatic equilibrium Earth; it rotates with the uniform velocity $\boldsymbol{\Omega}_0 = \Omega_0 \hat{e}_z$. It is thus different from the reference frame used in the torque approach.

3.10.10.1.2 Earth model parameters and unknowns used

Consider first a spherical Earth model. In this case, the physical parameters depend only on the radius r. The following are the relevant quantities:

- $\rho_0(r)$: the mean density on a sphere of radius r
- $\mu_0(r)$: the mean rigidity modulus on a sphere of radius r (it is 0 in the fluid part)
- $\kappa_0(r)$: the mean incompressibility modulus on a sphere of radius r
- $\lambda_0(r) = \kappa_0(r) - 2/3(\mu_0(r))$
- $\phi_0(r)$: mean initial self-gravity potential on a sphere of radius r

$\lambda_0(r)$ and $\mu_0(r)$ are called Lamé parameters; $\phi_0(r)$ is determined by the density function

$$\phi_0(r) = -G\int_{V_{\text{in}}} \frac{\rho_0(r')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}V_{\text{in}} \qquad [139]$$

where V_{in} is the volume included in the sphere of radius r. As the Earth in these theories is considered as being in equilibrium under rotation at a constant rate, it has an equatorial bulge that can be computed from the expression of the effect of the rotation at a long timescale (hydrostatic equilibrium). In such equilibrium, the equal rheological properties are constant on any self-gravity equipotential surface. For such an ellipsoidal rotating Earth of flattening $\epsilon(r)$, the preceding parameters depend not only on the radius but also on the colatitude θ, and one must add the centrifugal potential to the self-gravitational potential $\Phi_0(r) = \phi_0(r) + \psi_0(r)$. The centrifugal force related to the rotation $\boldsymbol{\Omega}_0$ is the negative gradient of the centrifugal potential, which is $-\frac{1}{2}\left[\boldsymbol{\Omega}_0^2 r^2 - (\boldsymbol{\Omega}_0 \cdot \mathbf{r})^2\right]$ It may be expressed as $-\nabla[\psi_0(r) + \psi_1(r) P_2(\cos(\theta))]$ where $\psi_0(r) = \frac{-\Omega^2 r^2}{3}$ at a distance r from the center of the Earth.

Let us take the density as an example. It can be written at the first order in the flattening:

$$\rho(r, \theta) = \rho_0(r) + \rho_2(r) P_2(\cos(\theta)) \qquad [140]$$

where

$$\rho_2(r) = \frac{2}{3}\varepsilon(r_0) r_0 \frac{\mathrm{d}\rho_0}{\mathrm{d}r} \qquad [141]$$

and P_2 is the Legendre polynomial of degree 2. One can write down similar expressions for the increments to the Lamé parameters $\lambda_0(r), \mu_0(r)$ and to $\Phi_0(r)$ in terms of $\mu_2(r), \lambda_2(r)$ and $\Phi_2(r)$.

The parameters in the preceding text are given as input to the deformation equations that will be integrated to obtain the nutations. The unknowns of our system are (1) the components of the vector field $s(\mathbf{r}, t)$ representing the displacements of the points $\mathbf{r}$ in the Earth caused by the gravitational action of the external body, (2) the components of the stress tensor $\mathbf{T}(\mathbf{r}, t)$ (or more precisely in this context, the components of $\mathbf{T}(\mathbf{r}, t) \bullet \hat{e}_r$, the stress components on a surface element normal to the radius vector), and (3) the mass redistribution potential Φ_1^{E} induced by the deformational response to the external potential.

3.10.10.1.3 Basic equations

3.10.10.1.3.1 Poisson equation

The Poisson equation relates the mass redistribution potential to the displacement field, more specifically to the dilatation and the density gradient:

$$\nabla^2 \Phi_1^{\mathrm{E}} = -4\pi G \nabla\cdot[(\rho_0 + \rho_2 P_2)\mathrm{s}] \qquad [142]$$

3.10.10.1.3.2 Stress–strain relationship

The stress–strain relationship expresses the incremental stress tensor in terms of the deformation:

$$\mathbf{T} = (\lambda_0 + \lambda_2 P_2)(\nabla\cdot\mathrm{s})\mathbf{I} + (\mu_0 + \mu_2 P_2)\left(\nabla\cdot\mathrm{s} + (\nabla\cdot\mathrm{s})^{\mathrm{T}}\right) \qquad [143]$$

where $\mathbf{I}$ is the unit matrix and where superscript T indicates that we must take the transpose of the matrix.

3.10.10.1.3.3 Motion equation

This is the equation relating the forces acting on a body and its deformations. It involves the internal stress, the self-gravity, the rotation of the reference frame, the external gravitational potential $\boldsymbol{\Phi}^{\mathrm{ext}}$, and the mass redistribution potential:

$$\nabla\cdot\mathbf{T} = (\rho_0 + \rho_2 P_2)(\eta_1 + \eta_2 + \eta_3) \qquad [144]$$

where the η_i are defined by

$$\eta_1 = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2} + \nabla\Phi_1^{\mathrm{E}} + \nabla\Phi^{\mathrm{ext}} - (\nabla\cdot s)\nabla\Phi_0 + \nabla(s\cdot\nabla\Phi_0) \qquad [145]$$

$$\eta_2 = 2\boldsymbol{\Omega} \wedge \frac{\mathrm{d}s}{\mathrm{d}t} \qquad [146]$$

$$\eta_3 = -(\nabla\cdot s)\nabla(\Phi_2 P_2) + \nabla(\mathrm{s}\cdot\nabla(\Phi_2 P_2)) \qquad [147]$$

3.10.10.1.3.4 Boundary conditions

The quantities appearing in the previous set of equations have to obey boundary conditions appropriate to the boundaries involved. With the use of the generic symbol $\hat{\mathbf{n}}$ for the unit normal to any boundary surface, the boundary conditions are the following:

1. Continuity of $\hat{\mathbf{n}}\cdot\mathbf{s}, \hat{\mathbf{n}}\cdot\mathbf{T}, \Phi_1$ and $\hat{\mathbf{n}}\cdot(\nabla\Phi_1 + 4\pi G\rho s)$ across every boundary, with $\hat{\mathbf{n}}\cdot\mathbf{T}$ required to vanish at the outer (free) surface, assuming that there is no surface loading.
2. Continuity of s across any welded boundary between two solid regions.
3. Conditions at fluid–solid boundaries: Inviscid fluids cannot support shear stresses. Consequently, the part of $\hat{\mathbf{n}}\cdot\mathbf{T}$ that is perpendicular to $\hat{\mathbf{n}}$ vanishes on the fluid side of the boundary, so it must vanish on the solid side too, in view of the previously stated continuity condition on $\hat{\mathbf{n}}\cdot\mathbf{T}$. Similarly, if the core fluid has no viscosity, it can 'glide' along the solid boundary, making it possible for the components of s that are tangential to the boundary to be discontinuous across the boundary.
4. Conditions at $r = 0$: Except for nonzero stress at the center of the Earth, all the variables ($\boldsymbol{\Phi}_1$, $\nabla\boldsymbol{\Phi}_1$, and s) vanish at the center.

3.10.10.1.4 Generalized spherical harmonics (GSH) expansions: radial functions

Solution of the equations of the problem is facilitated by reducing the equations to those for individual spectral components of the variables governed by the equations. For a spectral component of frequency σ cpsd, the time dependence appears through the factor $\mathrm{e}^{i\sigma\Omega_0 t}$, and therefore, the time derivative $\partial/\partial t$ appearing in eqns [145] and [146] may be replaced by $i\sigma\Omega_0$. The fundamental eqns [142]–[144] then become partial differential equations in the spatial variables only.

As a second step, one projects these equations on to a basis $(\hat{e}_-, \hat{e}_0, \hat{e}_+)$ related to the classical basis vectors $(\hat{e}_r, \hat{e}_\theta, \hat{e}_\lambda)$ for a reference frame appropriate for spherical coordinates. The unit vectors $(\hat{e}_-, \hat{e}_0, \hat{e}_+)$ constitute the canonical basis defined by

$$\begin{aligned} \hat{e}_- &= \frac{1}{\sqrt{2}}(\hat{e}_\theta - \mathrm{i}\hat{e}_\lambda) \\ \hat{e}_0 &= \hat{e}_r \\ \hat{e}_+ &= -\frac{1}{\sqrt{2}}(\hat{e}_\theta + \mathrm{i}\hat{e}_\lambda) \end{aligned} \qquad [148]$$

This basis is used in conjunction with the GSH D^l_{mn}, which are functions of the colatitude θ and the longitude ϕ and which depend on three integer numbers: l in the interval $[0, +\infty]$, called the degree, m in $[-l, +l]$, called the order, and n, determined by the tensor quantity considered in $[-l, +l]$ (n is in $[-1, +1]$ for a vector and in $[-2, +2]$ for a tensor, etc.) (see also **Annex B** where canonical decompositions of vector, scalar, and tensor fields are illustrated). Then, the coefficients of any particular GSH in the expansions of the fields s, Φ_1^{E}, etc., and hence in the various terms of the earlier mentioned partial differential equations, are functions of r only; when the expansions are used, these equations reduce to coupled ordinary differential equations for the radial functions. For the displacement field s, for instance, one writes the expansion

$$\begin{aligned} \mathrm{s}(r,\theta,\lambda) &= s^-\hat{e}_- + s^0\hat{e}_0 + s^+\hat{e}_+ = \begin{pmatrix} s_- \\ s_0 \\ s_+ \end{pmatrix} \\ &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \begin{pmatrix} S_l^{m_-}(r)D^l_{m_-}(\theta,\lambda) \\ S_l^{m_0}(r)D^l_{m_0}(\theta,\lambda) \\ S_l^{m_+}(r)D^l_{m_+}(\theta,\lambda) \end{pmatrix} \end{aligned} \qquad [149]$$

We may reexpress eqn [149] in a form which makes explicit the fact that the vector field s is made up of three geometrically distinct parts. For this purpose, define $U_l^m(r), V_l^m(r), W_l^m(r)$ by

$$U_\ell^m(r) = S_\ell^{m0}(r) \qquad [150]$$

$$V_\ell^m(r) = S_\ell^{m_+}(r) + S_\ell^{m_-}(r) \qquad [151]$$

$$W_\ell^m(r) = S_\ell^{m+}(r) - S_\ell^{m-}(r) \qquad [152]$$

Similarly, the decomposition of the radial tension is given by

$$\mathbf{T}\cdot\hat{e}_r = \sum_{l=0}^{\infty}\sum_{m=-l}^{\infty} \begin{pmatrix} T_{l0}^{m-}D^l_{m-} \\ T_{l0}^{m0}D^l_{m0} \\ T_{l0}^{m+}D^l_{m+} \end{pmatrix} \qquad [153]$$

In terms of these quantities, the expansion of s goes over into

$$\mathrm{s}(r,\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}U_l^m D_{m0}^l(\theta,\phi)\hat{e}_0 + \frac{1}{2}V_l^m\left[D_{m-}^l(\theta,\phi)\hat{e}_+ + D_{m+}^l(\theta,\phi)\hat{e}_-\right] + \frac{1}{2}W_l^m\left[D_{m-}^l(\theta,\phi)\hat{e}_+ - D_{m+}^l(\theta,\phi)\hat{e}_-\right] \quad [154]$$

We observe that the first of the three terms in the preceding text is a radial vector field, since $\hat{e}_0$ is simply the unit vector $\hat{\mathbf{r}}$ in the radial direction. The second and third terms are transverse to the radial direction; the former (like the radial part) is deformational and is called the spheroidal part of $s(r, \boldsymbol{\theta}, \boldsymbol{\phi},)$, while the latter is purely rotational and is called the toroidal part. (The two deformational parts may more precisely be designated as the radial and transverse spheroidal fields.) Any other vector field may also be similarly decomposed to reflect the geometric characters of the three different parts. For instance, starting with the GSH expansion

$$\mathbf{T}\cdot\hat{e}_r=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\begin{pmatrix}T_{l0}^{m-}D_{m-}^l\\ T_{l0}^{m0}D_{m0}^l\\ T_{l0}^{m+}D_{m+}^l\end{pmatrix} \quad [155]$$

of the vector field $\mathbf{T}\bullet\hat{e}_r$ representing the radial component of the stress tensor, one may reduce it to the same form as eqn [154], with (U_l^m, V_l^m, W_l^m) replaced by (P_l^m, Q_l^m, R_l^m) where

$$P_l^m(r)=T_{l0}^{m0}(r) \quad [156]$$

$$Q_l^m(r)=T_{l0}^{m+}(r)+T_{l0}^{m-}(r) \quad [157]$$

$$R_l^m(r)=T_{l0}^{m+}(r)-T_{l0}^{m-}-(r) \quad (158)$$

Besides the two vector fields in the preceding text, the scalar field $\Phi_1^{\mathrm{E}}(r,\theta,\lambda)$ too appears in the differential equations of our problem. The GSH expansion of this field

$$\Phi_1^{\mathrm{E}}(r,\theta,\lambda)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\Phi_{1l}^{\mathrm{Em}}(r)D_{m0}^l(\theta,\lambda) \quad [159]$$

introduces a new radial function $\Phi_{1l}^{\mathrm{Em}}(r)$. It is convenient to introduce another radial function

$$g_{1l}^{\mathrm{Em}}(r)=\frac{\mathrm{d}\Phi_{1l}^{\mathrm{Em}}(r)}{\mathrm{d}r}+4\pi G\rho_0 U_l^m(r) \quad (160)$$

From now on, we shall omit to write the dependence on r, for simplifying the writing.

The system of equations for our problem may now be expressed as a set of equations for the eight radial functions $U_l^m(r),\ldots,R_l^m(r),\Phi_{1l}^{\mathrm{Em}}(r),g_{1l}^{\mathrm{Em}}(r)$.

3.10.10.1.5 Equations for radial functions

3.10.10.1.5.1 Poisson equation

The projection of the Poisson equation on the GSH basis provides one with an equation of the second order in d/dr. The introduction of the new variable g_1^{E} enables us to reduce this second-order equation to a pair of first-order equations:

$$\frac{\mathrm{d}\Phi_{1l}^{\mathrm{Em}}}{\mathrm{d}r}=g_{1l}^{\mathrm{Em}}-4\pi G\rho_0 U_l^m \quad [161]$$

$$\frac{\mathrm{d}g_{1l}^{\mathrm{Em}}}{\mathrm{d}r}=\frac{l(l+1)}{r^2}\Phi_{1l}^{\mathrm{Em}}-\frac{2}{r}g_{1l}^{\mathrm{Em}}-\frac{L_0^l}{r}4\pi G\rho_0 V_l^m -4\pi G\sum_{l'=l-2}^{|l+2|}\begin{vmatrix}l&2&l'\\1&0&1\\m&0&m\end{vmatrix}\frac{L_0^l}{r}\rho_2\begin{Bmatrix}V_{l'}^m\\W_{l'}^m\end{Bmatrix} +4\pi G\sum_{l'=l-2}^{|l+2|}\begin{vmatrix}l&2&l'\\0&0&0\\m&0&m\end{vmatrix}\left[\frac{\mathrm{d}\rho_2}{\mathrm{d}r}U_{l'}^m+\frac{2}{r}\rho_2 U_{l'}^m +\rho_2\left(\frac{1}{\beta_0}P_{l'}^m-\frac{\lambda_0}{\beta_0 r}\left(L_0^{l'}V_{l'}^m+2U_{l'}^m\right)\right)\right] \quad [162]$$

where $\beta_0=\lambda_0+2\mu_0$ and $\beta_2=\lambda_2+2\mu_2$ and L_0^l and $L_0^{l'}$ are defined by

$$L_0^l=\sqrt{\frac{l(l+1)}{2}} \quad [163]$$

$$L_0^{l'}=\sqrt{\frac{l'(l'+1)}{2}} \quad [164]$$

The last factor in the second term of eqn [162] consists of two elements written as a column enclosed by curly brackets. The column is meant to be equal to its top element if l and l' have the same parity and to the bottom element if l and l' are of opposite parity. The same shorthand notation may be found in eqns [166] and [167] and in a number of equations further in the succeeding text.

3.10.10.1.5.2 Stress–strain relationship

The projection of the stress–strain relationship on the GSH basis provides three differential equations of the first order in d/dr:

$$\frac{\mathrm{d}U_l^m}{\mathrm{d}r}=\frac{1}{\beta_0}P_l^m-\frac{\lambda_0}{\beta_0 r}\left(L_0^l V_l^m+2U_l^m\right)-\sum_{l'=l-2}^{|l+2|}\begin{vmatrix}l&2&l'\\0&0&0\\m&0&m\end{vmatrix}\times\left[\frac{\beta_2}{\beta_0}P_{l'}^m-2\frac{\lambda_0\mu_2-\lambda_2\mu_0}{\beta_0^2 r}\left(L_0^{l'}V_{l'}^m+2U_{l'}^m\right)\right] \quad [165]$$

$$\frac{\mathrm{d}V_l^m}{\mathrm{d}r}=\frac{1}{\mu_0}Q_l^m-\frac{1}{r}\left(V_l^m+2L_0^l U_l^m\right)-\frac{\mu_2}{\mu_0}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix}l&2&l'\\1&0&1\\m&0&m\end{vmatrix}\begin{Bmatrix}Q_{l'}^m\\R_{l'}^m\end{Bmatrix} \quad [166]$$

$$\frac{\mathrm{d}W_l^m}{\mathrm{d}r}=\frac{1}{\mu_0}R_l^m+\frac{1}{r}W_l^m-\frac{\mu_2}{\mu_0}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix}l&2&l'\\1&0&1\\m&0&m\end{vmatrix}\begin{Bmatrix}R_{l'}^m\\Q_{l'}^m\end{Bmatrix} \quad [167]$$

The last factor in each of the last two equations consists of two elements written as a column enclosed by curly brackets. The column is meant to be equal to its top element if l and l' have the same parity and to the bottom element if l and l' are of opposite parity. The same shorthand notation may be found in a number of equations further in the succeeding text.

3.10.10.1.5.3 Motion equation

Finally, the projection of the motion equation on the GSH basis provides three differential equations of the first order in d/dr:

$$\frac{dP_l^m}{dr} = -\frac{2}{r}P_l^m - \frac{L_0^l}{r}Q_l^m + \frac{2}{r}\left[\lambda_0 X_l^m + \frac{\mu_0}{r}\left(L_0^l V_l^m + 2U_l^m\right)\right] + \rho_0\left(\eta_{1l}^{Um} + \eta_{2l}^{Um} + \eta_{3l}^{Um}\right) + \sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 0 & 0 & 0 \\ m & 0 & m \end{vmatrix} \left[\rho_2\left(\eta_{1l'}^{Um} + \eta_{2l'}^{Um} + \eta_{3l'}^{Um}\right) + \frac{2}{r}\left(\lambda_2 X_{l'}^m + \frac{\mu_2}{r}\left(L_0^{l'} V_{l'}^m + 2U_{l'}^m\right)\right)\right] \quad [168]$$

$$\frac{dQ_l^m}{dr} = -\frac{3}{r}Q_l^m - \frac{2\left(L_2^l\right)^2}{r^2}\mu_0 V_l^m + \frac{2L_0^l}{r}\left[\lambda_0 X_l^m + \frac{\mu_0}{r}\left(L_0^l V_l^m + 2U_l^m\right)\right] + \rho_0\left(\eta_{1l}^{Vm} + \eta_{2l}^{Vm} + \eta_{3l}^{Vm}\right) + \sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 1 & 0 & 1 \\ m & 0 & m \end{vmatrix}\rho_2\begin{Bmatrix} \eta_{1l'}^{Vm} + \eta_{2l'}^{Vm} \\ \eta_{1l'}^{Wm} + \eta_{2l'}^{Wm} \end{Bmatrix} + \frac{2L_0^l}{r}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 0 & 0 & 0 \\ m & 0 & m \end{vmatrix} \times \left[\lambda_2 X_{l'}^m + \frac{\mu_2}{r}\left(L_0^{l'} V_{l'}^m + 2U_{l'}^m\right)\right] + \frac{2L_2^l}{r^2}\mu_2\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 2 & 0 & 2 \\ m & 0 & m \end{vmatrix} L_2^{l'}\begin{Bmatrix} V_{l'}^m \\ W_{l'}^m \end{Bmatrix} \quad [169]$$

$$\frac{dR_l^m}{dr} = -\frac{3}{r}R_l^m + \frac{2\left(L_2^l\right)^2}{r^2}\mu_0 W_l^m + \rho_0\left(\eta_{1l}^{Wm} + \eta_{2l}^{Wm} + \eta_{3l}^{Wm}\right) + \sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 1 & 0 & 1 \\ m & 0 & m \end{vmatrix}\rho_2\begin{Bmatrix} \eta_{1l'}^{Wm} + \eta_{2l'}^{Wm} \\ \eta_{1l'}^{Vm} + \eta_{2l'}^{Vm} \end{Bmatrix} + \frac{2L_2^l}{r^2}\mu_2\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 2 & 0 & 2 \\ m & 0 & m \end{vmatrix} L_2^{l'}\begin{Bmatrix} W_{l'}^m \\ V_{l'}^m \end{Bmatrix} \quad [170]$$

where L_0^l, $L_0^{l'}$ and L_2^l were defined previously by eqns [163] and [164], and $L_2^{l'}$ are defined by

$$L_2^l = \sqrt{\frac{(l-1)(l+2)}{2}} \quad [171]$$

$$L_2^{l'} = \sqrt{\frac{(l'-1)(l'+2)}{2}} \quad [172]$$

Note that X_l^m appearing in eqns [168] and [169] are defined by

$$X_l^m = \frac{dU_l^m}{dr} + \frac{1}{r}\left(L_0^l V_l^m + 2U_l^m\right) \quad [173]$$

The radial, spheroidal, and toroidal scalars $\eta_{il'}^{Um}$, $\eta_{il'}^{Vm}$, and $\eta_{il'}^{Wm}$ are defined by

$$\eta_{1l}^{Um} = -\Omega^2\sigma^2 U_l^m + g_{1l}^{\mathrm{Em}} - \frac{2(g+\tilde{g})}{r}U_l^m - \frac{L_0^l}{r}\tilde{g}V_l^m \quad [174]$$

$$\eta_{1l}^{Vm} = -\Omega^2\sigma^2 V_l^m - \frac{L_0^l}{r}\left(\Phi_{1l}^{\mathrm{Em}} + \tilde{g}U_l^m\right) \quad [175]$$

$$\eta_{1l}^{Wm} = -\Omega^2\sigma^2 W_l^m \quad [176]$$

$$\eta_{2l}^{Um} = 2\sigma\Omega^2\sum_{l'=|l-1|}^{|l+1|}\begin{vmatrix} l & 2 & l' \\ 0 & 1 & -1 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} V_{l'}^m \\ -W_{l'}^m \end{Bmatrix} \quad [177]$$

$$\eta_{2l}^{Vm} = -2\sigma\Omega^2\sum_{l'=|l-1|}^{|l+1|}\left[\begin{vmatrix} l & 2 & l' \\ -1 & -1 & 0 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} 2U_{l'}^m \\ 0 \end{Bmatrix} + \begin{vmatrix} l & 2 & l' \\ -1 & 0 & -1 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} V_{l'}^m \\ -W_{l'}^m \end{Bmatrix}\right] \quad [178]$$

$$\eta_{2l}^{Wm} = 2\sigma\Omega^2\sum_{l'=|l-1|}^{|l+1|}\left[\begin{vmatrix} l & 2 & l' \\ -1 & -1 & 0 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} 0 \\ 2U_{l'}^m \end{Bmatrix} + \begin{vmatrix} l & 2 & l' \\ -1 & 0 & -1 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} -W_{l'}^m \\ V_{l'}^m \end{Bmatrix}\right] \quad [179]$$

$$\eta_{3l}^{Um} = \sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 0 & 0 & 0 \\ m & 0 & m \end{vmatrix} \times \left[\frac{d\Phi_2}{dr}\left(\frac{dU_{l'}^m}{dr} - X_{l'}^m\right) + \frac{d^2\Phi_2}{dr^2}U_{l'}^m\right] + \sqrt{3}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 0 & 1 & -1 \\ m & 0 & m \end{vmatrix} \times \left(\frac{1}{r}\frac{d\Phi_2}{dr} - \frac{\Phi_2}{r^2} - \frac{\Phi_2}{r}\frac{d}{dr}\right)\begin{Bmatrix} V_{l'}^m \\ -W_{l'}^m \end{Bmatrix} \quad [180]$$

$$\eta_{3l}^{Vm} = -\frac{2L_0^l}{r}\frac{d\Phi_2}{dr}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 0 & 0 & 0 \\ m & 0 & m \end{vmatrix}U_{l'}^m - \sqrt{3}\frac{\Phi_2}{r}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 0 & 1 & -1 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} V_{l'}^m \\ -W_{l'}^m \end{Bmatrix} + 2\sqrt{3}\frac{\Phi_2}{r}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 1 & 1 & 0 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} X_{l'}^m \\ 0 \end{Bmatrix} \quad [181]$$

$$\eta_{3l}^{Wm} = 2\sqrt{3}\frac{\Phi_2}{r}\sum_{l'=l-2}^{|l+2|}\begin{vmatrix} l & 2 & l' \\ 1 & 1 & 0 \\ m & 0 & m \end{vmatrix}\begin{Bmatrix} 0 \\ X_{l'}^m \end{Bmatrix} \quad [182]$$

where the superscripts U, V, and W indicate that the quantities $\eta_{il}^{U,V,W,m}$ appear in the equation for dU_l^m/dr, dV_l^m/dr, and dW_l^m/dr, respectively. $\tilde{g}$ and g appearing in these equations are defined by $\tilde{g} = \frac{dW_0}{dr}$ and $g = 2\pi G\rho_0 r + \frac{r}{2}\frac{d^2W_0}{dr}$.

It is to be noted that σ, the frequency of excitation of the deformation, enters the equations of motion through the expressions in eqns [174]–[179]. It is important to note that the radial differential equations obtained previously do not couple field components of different orders m. Furthermore, the radial functions that are coefficients of the spheroidal (including radial) parts of degree l in the GSH expansions are coupled with coefficients of toroidal parts of degrees $l-1$, $l+1$, $l-3$, $l+3\ldots$ and with coefficients of spheroidal parts of degrees $l-2$, $l+2$, $l-4$, $l+4\ldots$.

3.10.10.1.6 Solutions of the radial equations

Let us now arrange the set of radial functions contained in the spheroidal vector fields as a column vector $\bar{\sigma}_l^m$ and those in the toroidal fields as a column vector $\bar{\tau}_l^m$. In view of the nature of the couplings of the different parts of the field as just explained, the differential equations for the $\bar{\sigma}_l^m$ and $\bar{\tau}_l^m$ have the structure

of the following matrix-differential equation, wherein ε is a generic symbol standing for the matrix blocks made up of the first-order parts of the matrix elements that arise from the ellipticity of the Earth's structure and A_0 is for the zeroth-order parts, which remain nonvanishing when ellipticity is neglected. (The dimensions of the matrix blocks A_0 and the values of the elements of the blocks depend on their positions in the matrix (eqn [183]) and similarly for the blocks ε. All the blocks are rectangular.)

$$\begin{pmatrix} \cdot & & & & & & & & \\ & \cdot & & & & & & & \\ \varepsilon & \varepsilon & A_0+\varepsilon & \varepsilon & \varepsilon & 0 & 0 & 0 & 0 \\ 0 & \varepsilon & \varepsilon & A_0+\varepsilon & \varepsilon & \varepsilon & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & \varepsilon & A_0+\varepsilon & \varepsilon & \varepsilon & 0 & 0 \\ 0 & 0 & 0 & \varepsilon & \varepsilon & A_0+\varepsilon & \varepsilon & \varepsilon & 0 \\ 0 & 0 & 0 & 0 & \varepsilon & \varepsilon & A_0+\varepsilon & \varepsilon & \varepsilon \\ & & & & & & & \cdot & \\ & & & & & & & & \cdot \end{pmatrix} \times \begin{pmatrix} \cdot \\ \cdot \\ \bar{\sigma}_{l-2}^m \\ \bar{\tau}_{l-1}^m \\ \bar{\sigma}_l^m \\ \bar{\tau}_{l+1}^m \\ \bar{\sigma}_{l+2}^m \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ f(\Phi^{\text{ext}}) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad [183]$$

where $f(\Phi^{\text{ext}})$ represents a normalized function of the forcing potential Φ^{ext}, which is taken to be of degree l and order m, and $\bar{\sigma}_l^m$ and $\bar{\tau}_l^m$ stand for the following column vectors:

$$\bar{\sigma}_l^m = \begin{pmatrix} U_l^m \\ P_l^m \\ \Phi_{1l}^{\text{Em}} \\ g_{1l}^{\text{Em}} \\ V_l^m \\ Q_l^m \end{pmatrix} \quad [184]$$

and

$$\bar{\tau}_l^m = \begin{pmatrix} W_l^m \\ R_l^m \end{pmatrix} \quad [185]$$

The system of equations to be solved is infinite since l may take any integer value $\geq |m|$.

In order to solve the system, it is necessary to truncate at the first order in the small quantities, considering that $\bar{\sigma}_{l+2}^m$ and $\bar{\sigma}_{l-2}^m$ are small as well. With the degree l and order m fixed, one considers the following supertruncated system:

$$\begin{pmatrix} A_0+\varepsilon & \varepsilon & \varepsilon \\ \varepsilon & A_0+\varepsilon & \varepsilon \\ \varepsilon & \varepsilon & A_0+\varepsilon \end{pmatrix} \begin{pmatrix} \bar{\tau}_{l-1}^m \\ \bar{\sigma}_l^m \\ \bar{\tau}_{l+1}^m \end{pmatrix} = \begin{pmatrix} 0 \\ f(\Phi^{\text{ext}}) \\ 0 \end{pmatrix} \quad [186]$$

consisting in only the ten ordinary differential equations in d/dr, involving the following ten variables: $U_l^m, P_l^m, \Phi_{1l}^{\text{Em}}, g_{1l}^{\text{Em}}, V_l^m, Q_l^m, W_{l-1}^m, R_{l-1}^m, W_{l+1}^m$, and R_{l+1}^m.

For the actual solution of these equations, one starts then from five independent solutions that do not diverge at $r=0$. One integrates from the center up to the surface of the Earth using as inputs the rheological properties λ_0 and μ_0 as functions of r, as well as the density function ρ_0, and using the boundary conditions at the ICB and at the CMB (such as continuity of the potential, continuity of the radial displacement, and continuity of the stress). One reaches the surface with five independent solutions, to which one applies the five surface boundary conditions. This ensures the uniqueness of the solution and is done for one given frequency.

If one wants to search for the Earth's normal modes, one has to set $\Phi^{\text{ext}}=0$ in the truncated equations, integrate similarly inside the Earth (ignoring the right-hand side of eqn [186]), and impose the boundary conditions on the solutions of the resulting system of equations. At the surface, the boundary conditions to be satisfied correspond to algebraic relations that have a solution only when the determinant of the matrix of the system equals zero, which is a relation involving the frequency. The remaining steps are then to search by successive iterations for the frequency that allows the system to be solved (that corresponds to a determinant of the matrix equal to zero).

3.10.11 Atmospheric Tides and Nontidal Effects from Surficial Fluids

The effects of fluid masses at the Earth's surface on nutation must be considered as they produce observable effects. They consist of ocean tidal effects, atmospheric effects, and nontidal ocean effects. The first of these has been dealt with in Section 3.10.9.4.

Atmospheric effects on nutation are mainly on the prograde annual component corresponding to the one solar day period in a terrestrial frame; the diurnal cycle in the temperature of the atmosphere is responsible for the effect. Modulation of the diurnal frequency by the annual and semiannual seasonal variations results in effects at additional frequencies, which translate into semiannual and terannual prograde frequencies and the retrograde annual and zero frequencies in the celestial frame. These effects are not easy to compute, and in any case, the amplitudes of the atmospheric signals keep changing with time. Further consideration of the atmospheric effects on nutation should be made in the time domain. Nevertheless, a new estimation of the atmospheric contributions to Earth's nutations based on three reanalyses of atmospheric global circulation models (GCM) has been recently used by Koot and de Viron (2011) in order to estimate the complex amplitudes of the periodic terms in the atmospheric forcing. Unlike previous estimations based on operational GCMs, the results obtained from the three reanalysis GCMs are in good agreement, and it was possible to show that the atmospheric contributions to nutations are quite small. The implications of the atmospheric contribution to the estimation of Earth's deep interior properties from nutation observations have been studied as well. Koot and de Viron (2011) showed that this contribution is too small to affect significantly the estimation of these properties.

Atmospheric effects on the ocean also induce changes in the nutation. These too are difficult to compute. In practice, the annual prograde nutation contribution from both direct atmospheric effect and induced atmospheric ocean effect on

nutation is fitted on the observation. de Viron et al. (2005) have shown that the effects computed from general circulation models using the angular momentum approach fit the 'observed' value surprisingly well.

3.10.12 New Conventions for Earth Rotation Variations

3.10.12.1 CIP and Relation of its Motions to the Nutation and Wobble

The CIP is a conceptual pole (see Capitaine, 2000; Capitaine et al., 2000), the motion of which corresponds to the precessional motion, and nutational motions with periods greater than 2 days, of the ITRF pole in space. Its position is represented by two coordinates (X, Y) in the International Celestial Reference Frame (ICRF), which include a constant offset from the pole of the ICRF besides the lunisolar precessions and nutations in longitude and obliquity. Reduction of observational data requires the use of the transformation between the ICRF and the ITRF, which involves a succession of elementary rotations (represented by matrices), which may be chosen in different ways. One of the alternative expressions is in terms of the rotation, which carries the pole of the ICRF into the intermediate pole (CIP) and the ICRF into an intermediate reference frame, a rotation of this frame around the CIP, representing the axial rotation of the Earth, and a further rotation of the resulting reference frame to bring it into coincidence with the ITRF and the CIP to the pole of the ITRF.

Precession and nutation, as stated previously, are defined now by the motion of the CIP relative to the ICRF, and wobble is the motion of the CIP relative to the ITRF. The rotation around the CIP will be dealt with in more detail in the next section.

3.10.12.2 Nonrotating Origin

Time based on Earth rotation, as classically defined, is the Greenwich Apparent Sidereal Time (GAST, or simply Greenwich Sidereal Time (GST)). GST is the arc length measured along the equator of date from the equinox of date to the Greenwich meridian, which marks the axis Ox that serves as the origin of longitude. If the axis of rotation (the pole of the equator) is kept in a fixed direction in space (and if the ecliptic remained stationary), GST thus defined would give the angle of Earth rotation exactly. However, the equatorial plane is moving in space because of nutation and precession, and this gives rise to small motions of the equinox along the equator that mimic small rotations around the pole of the equator, though they are not caused by axial rotation around the pole of the equator. GST is therefore no longer an exact measure of the angle of axial rotation. In order to obtain a rigorous measure of time based on Earth rotation, it is necessary, therefore, to define an origin on the equator, which has no motion along the instantaneous equator or, differently stated, which has no rotation around the pole of the equator. Such an origin is referred to as the nonrotating origin (NRO) (Guinot, 1979).

According to the conventions adopted by the IAU in 2000 and the IUGG in 2003, the axial rotation of the Earth is to be measured around the CIP, so the equator of the CIP is the relevant equator for the NRO, which is to replace the equinox in the definition of time. To distinguish the newly defined NRO from other origins, which had earlier been referred to by the same name, the nomenclature CIO (Celestial Intermediate Origin) and TIO have been adopted for the NROs on the CIP equator, by IAU2000 resolutions (see Capitaine, 1986, 1990; Capitaine et al., 1986).

If $\hat{n}$ is the unit vector in the direction of the CIP and if Ω is the angular velocity of the Earth, the defining condition on the NRO, namely, that it should have no rotation around the CIP, is given by (see Capitaine, 1986, 1990; Capitaine et al., 1986)

$$\hat{n}\cdot\Omega = 0 \qquad [187]$$

An equivalent way of stating the condition is in terms of the velocity of the CIO – that the velocity, say $\dot{x}$, should be parallel to the instantaneous direction of the CIP:

$$\dot{x} = k\hat{n} \qquad [188]$$

On taking the scalar product with $\hat{n}$, one finds that $k = x\cdot\hat{n}$, the last step being a consequence of the fact that $x\cdot\hat{n}$ and its time derivative vanish since x is on the equator of $\hat{n}$. Hence, the defining condition takes the form introduced by Kaplan (2005):

$$\dot{x} = -(x\cdot\hat{n})\hat{n} \qquad [189]$$

Considering that $\dot{x} = \Omega \wedge x$, and that consequently $\dot{x}\cdot\Omega = 0$, one immediately sees that eqns [188] and [189] are perfectly equivalent.

3.10.12.3 Definition of Universal Time

Until the implementation of the IAU2000 resolutions, UT1 was defined from GMST through the relation

$$\mathrm{GMST}(T_u, \mathrm{UT1}) = \mathrm{GMST}_{0h\mathrm{UT1}}(T_u) + r\mathrm{UT1}$$

which relates the increase of GMST during day T_u to the 'time of day' (UT1), with $\mathrm{GMST}_{0h\mathrm{UT1}}(T_u)$ given by a formula that is not strictly linear in T_u (see IERS Conventions, 1996; Capitaine et al., 1986, 2000, 2003; McCarthy, 1996).

Universal time UT1 was intended to be a measure of the true accumulated rotation of the Earth after a specified initial epoch. The stellar angle is defined to be just that. GMST does not represent the true magnitude of the accumulated rotation. The new definition of UT1 is in terms of $\theta : \theta = \theta_0 + k(\mathrm{UT1} - \mathrm{UT1}_0)$, where k is a scale factor that accounts for the fact that the rate of change of θ differs slightly from that of the old UT1. In this case, the new day of UT1 remains close to the mean solar day. The value of θ_0 is chosen to ensure continuity of the new UT1 with the old one at the epoch of switchover from the old to the new (2003 January 1.0, Julian Day 2452640.5).

3.10.12.4 Transformation between ICRF and ITRF

Reduction of observational data requires the use of the transformation between the ICRF and the ITRF, which involves a succession of elementary rotations (represented by matrices) that may be chosen in different ways as stated previously. They may, for instance, be rotations through the classical Euler angles. An alternative expression is in terms of the rotation that carries the pole of the ICRF into an intermediate pole, a

rotation around this pole, and the rotation that brings this pole into coincidence with the pole of the ITRF.

Classically, the matrix transformation that allows us to pass from the TRF to the CRF or vice versa was related to precession, nutation, LOD, and polar motion by

$$[\mathrm{TRF}] = WR_3(\mathrm{GST})PN[\mathrm{CRF}] \qquad [190]$$

or

$$[\mathrm{CRF}] = PN'R_3(-\mathrm{GST})W'[\mathrm{TRF}] \qquad [191]$$

where W contains the polar motion contribution, R_3 indicates a rotation around the third (z) axis, and PN contains the precession–nutation contributions, respectively (see Mueller, 1981).

GST is the Greenwich Sidereal Time of date (also called the Greenwich Apparent Sidereal Time, GAST), that is, the angle on the true equator of date between the true equinox of date and the Greenwich meridian, which is the origin of longitude in the terrestrial frame (direction of Ox). The transformation [190] is the classical one involving an intermediate frame associated with the true equator and equinox of date.

In the NRO approach, one has

$$[\mathrm{TRF}] = MR_3(\theta)M'[\mathrm{CRF}] \qquad [192]$$

where M and M' are related to the CIP coordinates in space and in the terrestrial frame, respectively. The transformation [192] is with reference to an intermediate frame based on the concept of the CIP and NRO (adopted by the IAU in 2000 and the IUGG in 2003). As explained earlier, the CIP position is represented by two coordinates (X, Y) in the ICRF. One has

$$[\mathrm{TRF}] = R_1\left(y_p\right)R_2\left(x_p\right)R_3(-s')R_3(-\theta) \times R_3(s)\begin{pmatrix} 1 & -XY/2 & X \\ -XY/2 & 1 & Y \\ -X & -Y & 1 \end{pmatrix}[\mathrm{CRF}] \qquad [193]$$

where R_1, R_2, and R_3 refer to rotations about the x-, y- and z-axes of the frame that one is in at the relevant stage of the transformation. Thus, in the decomposition of the transformation between the CRF and TRF into polar motion, axial rotation about the CIP, and nutation–precession, θ appears in combination with s and s' defined by Capitaine et al. (1986).

A Annex 1

A.1 Rotational Normal Mode Definitions

The existence of more than one rotational normal mode for the Earth is due to the presence of the core regions that are not rigidly attached to one another or to the mantle region and the consequent possibility of differential rotations between different regions.

In each of the rotational modes, the rotation axes of the three regions, as well as the symmetry axis of the SIC, are offset from the symmetry axis of the mantle by different angles. (The magnitude of the angular offset of the rotation axis of each region from the mantle symmetry axis is the amplitude of the wobble of that region.) All four axes rotate with a common frequency (the frequency of the normal mode) around the mantle symmetry axis. The rotational normal modes not only contribute directly to Earth rotation variations but also give rise to resonances in the forced nutations and wobbles when the frequency of forcing is close to one or the other of the normal mode eigenfrequencies. The four rotational modes are the following:

- The CW is a low-frequency mode with a period of about 432 days in a TRF. If observed, it is characterized by an offset of the direction of the axis of rotation of the mantle from the figure axis of the ellipsoidal Earth, accompanied by a motion of the rotation axis around the figure axis in the prograde sense, that is, in the same sense as that of the diurnal rotation of the Earth. In this mode, the rotation axes of the three regions are nearly coincident. This is the only rotational eigenmode that could exist if the Earth were wholly solid. (The period would of course be different in that case.) The counterpart of the CW for a rigid Earth is the Euler wobble with a period close to 300 days, which is determined solely by the Earth's dynamical ellipticity $(C-A)/A$, that is, the fractional difference between the moments of inertia about the polar and equatorial axes. In fact, the period, when expressed in cpsd, is just $(1/e)$.
- The FCN is a free mode related to the existence of a flattened fluid core inside the Earth. If observed, it is associated with an offset of the rotation axis of the fluid core from the mantle symmetry axis (i.e., the amplitude of the wobble of the FOC), far larger than that of the rotation axis of the mantle itself. The frequency of this mode is nearly diurnal as seen in a terrestrial frame; for this reason, the motion as seen from the terrestrial frame is referred to as the NDFW. The motion of the rotation axes is accompanied by a circular motion of the mantle symmetry axis in space, which is the FCN; it is retrograde, that is, in the opposite sense to that of the diurnal Earth rotation and has a period of about 430 days. Ellipticity of the core CMB is essential for the existence of the FCN mode. It causes the equatorial bulge of the fluid mass to impinge on the mantle at the CMB in the course of the differential rotation of the core and the mantle about noncoincident axes; the pressure of the fluid on the boundary and the inertial reaction on the fluid core itself sustain the differential rotation between the core and the mantle in the FCN mode.
- The FICN is a mode related to the existence of a flattened inner core inside the fluid core. If excited, the amplitude of the wobble of the SIC dominates over the wobble amplitudes of the other regions; the figure axis of the SIC too is offset by almost the same amount as the rotation axis of the SIC. The FICN has, like the FCN, a nearly diurnal retrograde period in a TRF. The motion of the axes in space is prograde in this mode, unlike in the case of the FCN. For this reason, the FICN is also referred to as the PFCN; when it is necessary to emphasize the distinction between this mode and the FCN, the latter is referred to as the retrograde free core nutation (RFCN). The period of the FICN in space is about 1000 days. As in the case of the FCN, the fluid pressure acting on the ellipsoidal boundary (the ICB in this case) plays an important role here; but an even larger role is played by the gravitational coupling of the SIC to the mantle due to the offset between the symmetry axes of the two ellipsoidal regions.

- The ICW is a mode also related to the existence of a flattened inner core inside the fluid core. If excited, the offset of the symmetry axis of the SIC from that of the mantle exceeds the offsets of the three rotation axes by several orders of magnitude. It has a long period (longer than the CW) in a terrestrial frame and is prograde. The ellipsoidal structure of the SIC and of the rest of the Earth and the density contrast between the SIC and the fluid at the ICB are essential elements in the generation of this normal mode.

B Annex 2

B.1 Generalized Spherical Harmonics

The GSH function D^l_{mn} is a function of θ and φ depending on three integer numbers l in $[0, +\infty]$, called degree; m in $[-l, +l]$, called order; and n in $[-l, +l]$. Its definition is the following:

$$D^l_{mn}(\theta,\varphi) = (-1)^{m+n} P^{mn}_l(\mu) e^{im\varphi} \qquad [194]$$

where $\mu = \cos\theta$ and P^{mn}_l is the generalized Legendre function:

$$P^{mn}_l(\mu) = \frac{(-1)^{l-n}}{2^l (l-n)!} \sqrt{\frac{(l-n)!(l+m)!}{(l+n)!(l-m)!}} \sqrt{\frac{1}{(1-\mu)^{m-n}(1+\mu)^{m+n}}}$$

$$\frac{d^{l-m}}{d\mu^{l-m}} \left((1-\mu)^{l-n} (1+\mu)^{l+n} \right) \qquad [195]$$

The product of two GSH is expressed using the so-called J-square coefficients related to the Wigner symbols as seen in the text (see Edmonds, 1960):

$$D^{l_1}_{m_1 n_1} D^{l_2}_{m_2 n_2} = \sum_{l=|l_2+l_1|}^{l=|l_2-l_1|} \begin{vmatrix} l & l_1 & l_2 \\ n & n_1 & n_2 \\ m & m_1 & m_2 \end{vmatrix} D^l_{mn} \qquad [196]$$

where $n = n_1 + n_2$ and $m = m_1 + m_2$. The GSH form a basis on which the components of a tensor, of a vector, or of a scalar can be decomposed. The canonical component (α_1, α_2) of a tensor $\mathbf{T}$ of order 2 can be written as

$$\mathbf{T}^{\alpha 1, \alpha 2}(r,\theta,\varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} T^{\alpha 1, \alpha 2}(r) D^l_{mn}(\theta,\varphi) \qquad [197]$$

where $n = \sum_{i=1}^{2} \alpha_i$. The canonical components of a vector

$$\mathrm{v}(r,\theta,\varphi) = v^-(r,\theta,\varphi)\hat{e}_- + v^0(r,\theta,\varphi)\hat{e}_0 + v^+(r,\theta,\varphi)\hat{e}+ = \begin{bmatrix} v^-(r,\theta,\varphi) \\ v^0(r,\theta,\varphi) \\ v^+(r,\theta,\varphi) \end{bmatrix} \qquad [198]$$

can be decomposed as

$$\mathrm{v}(r,\theta,\varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \begin{bmatrix} v^{m-}_l(r) D^l_{m-}(\theta,\varphi) \\ v^{m0}_l(r) D^l_{m0}(\theta,\varphi) \\ v^{m}_l + (r) D^l_{m+}(\theta,\varphi) \end{bmatrix} \qquad [199]$$

For a scalar function $f(r, \theta, \varphi)$, the decomposition in GSH is

$$f(r,\theta,\varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f^m_l(r) D^l_{m0}(\theta,\varphi) \qquad [200]$$

References

Alfè D, Kresse G, and Gillan MJ (2000) Structure and dynamics of liquid iron under Earth's core conditions. *Physical Review B* 61: 132–142.

Andoyer H (1923) *Cours de Mécanique Céleste*, vol. 1. Paris: Gauthier-Villar.

Bretagnon P and Francou G (1988) Planetary theories in rectangular and spherical variables: VSOP87 solutions. *Astronomy and Astrophysics* 202: 309–315.

Bretagnon P, Francou G, Rocher P, and Simon J-L (1998) SMART97: A new solution for the rotation of the rigid Earth. *Astronomy and Astrophysics* 329: 329–338.

Bretagnon P, Rocher P, and Simon JL (1997) Theory of the rotation of the rigid Earth. *Astronomy and Astrophysics* 319: 305–317.

Brumberg VA, Bretagnon P, and Francou G (1992) Analytical algorithms of relativistic reduction of astronomical observations. In: Capitaine N (ed.) *Journées Systèmes de Références, June 1991, Paris*, pp. 141–148.

Buffett BA (1996a) Gravitational oscillations in the length of Day. *Geophysical Research Letters* 23: 2279–2282.

Buffett BA (1996b) A mechanism for decade fluctuations in the length of day. *Geophysical Research Letters* 23: 3803–3806.

Buffett BA (2010) Chemical stratification at the top of Earth's core: Constraints from observations of nutations. *Earth and Planetary Science Letters* 296: 367–372.

Buffett BA and Christensen U (2007) Magnetic and viscous coupling at the core–mantle boundary: Inferences from observations of the Earth's nutations. *Geophysical Journal International* 171: 145–152.

Buffett BA, Mathews PM, and Herring TA (2002) Modeling of nutation and precession: Effects of electromagnetic coupling. *Journal of Geophysical Research* 107(B4). http://dx.doi.org/10.1029/2000JB000056.

Capitaine N (1986) The Earth rotation parameters: Conceptual and conventional definitions. *Astronomy and Astrophysics* 162: 323–329.

Capitaine N (1990) The celestial pole coordinates. *Celestial Mechanics and Dynamical Astronomy* 48: 127–143.

Capitaine N (2000) Definition of the celestial ephemeris pole and the celestial ephemeris origin. In: Johnson K, McCarthy D, Luzum B, and Kaplan G (eds.) *Towards Models and Constants for Sub-microarcsecond Astrometry. IAU Colloquium 180*, pp. 153–163. Washington: IAU Colloquium.

Capitaine N, Chapront J, Lambert S, and Wallace P (2003) Expressions for the Celestial Intermediate Pole and Celestial Ephemeris Origin consistent with IAU2000A precession-nutation model. *Astronomy and Astrophysics* 400: 1145–1154.

Capitaine N and Guinot B (1988) A non-rotating origin on the instantaneous equator. In: Babcock AK and Wilkins GA (eds.) *The Earth's Rotation and Reference Frames for Geodesy and Geodynamics*, pp. 33–38. IAU Publications.

Capitaine N, Guinot B, and McCarthy DD (2000) Definition of the celestial ephemeris origin and of UT1 in the International Celestial Reference Frame. *Astronomy and Astrophysics* 355: 398–405.

Capitaine N, Souchay J, and Guinot B (1986) A non-rotating origin on the instantaneous equator – Definition, properties and use. *Celestial Mechanics* 39(3): 283–307.

Cartwright DE and Taylor RJ (1971) New computations in the tide-generating potential. *Geophysical Journal of the Royal Astronomical Society* 23: 45–74.

Chao BF, Ray RD, Gipson JM, Egbert GD, and Ma C (1996) Diurnal/semidiurnal polar motion excited by oceanic tidal angular momentum. *Journal of Geophysical Research* 101(B9): 20151–20164.

Chapront-Touzé M and Chapront J (1983) The lunar ephemeris ELP-2000. *Astronomy and Astrophysics* 124(1): 50–62.

Chapront-Touzé M and Chapront J (1988) ELP2000-85: A semi analytical lunar ephemeris adequate for historical times. *Astronomy and Astrophysics* 190: 342–352.

de Viron O, Schwarzbaum G, Lott F, and Dehant V (2005) Diurnal and sub-diurnal effects of the atmosphere on the Earth rotation and geocenter motion. *Journal of Geophysical Research* 110(B11): B11404. http://dx.doi.org/10.1029/2005JB003761.

De Vries D and Wahr JM (1991) The effects of the solid inner core and nonhydrostatic structure on the Earth's forced nutations and Earth tides. *Journal of Geophysical Research* 96(B5): 8275–8293.

Defraigne P, Dehant V, and Wahr JM (1996) Internal loading of an homogeneous compressible Earth with phase boundaries. *Geophysical Journal International* 125: 173–192.

Dehant V (1987a) Tidal parameters for an inelastic Earth. *Physics of the Earth and Planetary Interiors* 49: 97–116.

Dehant V (1987b) Integration of the gravitational motion equations for an elliptical uniformly rotating Earth with an inelastic mantle. *Physics of the Earth and Planetary Interiors* 49: 242–258.

Dehant V (1990) On the nutations of a more realistic Earth's model. *Geophysical Journal International* 100: 477–483.

Dehant V, Hinderer J, Legros H, and Lefftz M (1993) Analytical approach to the computation of the Earth, the outer core and the inner core rotational motions. *Physics of the Earth and Planetary Interiors* 76: 259–282.
Deleplace B and Cardin P (2006) Visco-magnetic torque at the core mantle boundary. *Geophysical Journal International* 167(2): 557-566, http://dx.doi.org/10.1111/j.1365-246X.2006.03180.x.
Dziewonski AD and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.
Edmonds AR (1960) *Angular Momentum in Quantum Mechanics*, 2nd edn. Princeton: Princeton University Press.
Folgueira M, Souchay J, and Kinoshita H (1998a) Effects on the nutation of the non-zonal harmonics of third degree. *Celestial Mechanics* 69(4): 373–402.
Folgueira M, Souchay J, and Kinoshita H (1998b) Effects on the nutation of C4m and S4m harmonics. *Celestial Mechanics* 70(3): 147–157.
Fukushima T (1991) Geodesic nutation. *Astronomy and Astrophysics* 244: 11–12.
Guinot B (1979) Basic problems in the kinematics of the rotation of the Earth. In: McCarthy DD and Pilkington PD (eds.) *Time and the Earth's Rotation*, pp. 7–18. D. Reidel Publishing Company.
Gwinn CR, Herring TA, and Shapiro II (1986) Geodesy by radio interferometry: Studies of the forced nutations of the Earth, 2. Interpretation. *Journal of Geophysical Research* 91: 4755–4765.
Hartmann T and Soffel M (1994) The nutation of a rigid Earth model: Direct influences of the planets. *Astronomy Journal* 108: 1115–1120.
Hartmann T, Soffel M, and Ron C (1999) The geophysical approach towards the nutation of a rigid Earth. *Astronomy and Astrophysics Supplement Series* 134: 271–286.
Hartmann T and Wenzel H-G (1995a) Catalogue HW95 of the tide generating potential. *Bulletin d'Informations Marées Terrestres* 123: 9278–9301.
Hartmann T and Wenzel H-G (1995b) The HW95 tidal potential catalogue. *Geophysical Research Letters* 22(24): 9278–9301.
Herring TA, Gwinn CR, and Shapiro II (1986) Geodesy by radio interferometry: Studies of the forced nutations of the Earth, 1. Data analysis. *Journal of Geophysical Research* 91: 4745–4754.
Herring TA, Mathews PM, and Buffett B (2002) Modeling of nutation-precession of a non-rigid Earth with ocean and atmosphere. *Journal of Geophysical Research* 107(B4). http://dx.doi.org/10.1029/2001JB000165.
Hori GI (1966) Theory of general perturbations with unspecified canonical variables. *Publications of the Astronomical Society of Japan* 18(4): 287–296.
Kaplan G (2005) Another look at non-rotating origins. In: Engvold O (ed.), *Proceedings of the IAU 25 General Assembly, Sidney, Australia, Joint Discussion 16: The International Celestial Reference System, Maintenance and Future Realizations*, Highlights of Astronomy, vol. 13, Responsible Eds. for JD16 R. Gaume, D. McCarthy, and J. Souchay, pp. 196–199.
Kinoshita H (1977) Theory of the rotation of the rigid Earth. *Celestial Mechanics* 15: 277–326.
Kinoshita H and Souchay J (1990) The theory of the nutation for the rigid Earth model at the second order. *Celestial Mechanics* 48: 187–265.
Klioner S (1998) Astronomical reference frames in the PPN formalism. In: Vondrak J and Capitaine N (eds.) *Journées Systèmes de Références, September 1997, Prague, Czech Republic,* pp. 32–37.
Koot L and de Viron O (2011) Atmospheric contributions to nutations and implications for the estimation of deep Earths properties from nutation observations. *Geophysical Journal International* 185: 1255–1265. http://dx.doi.org/10.1111/j.1365-246X.2011.05026.x.
Koot L and Dumberry M (2011) Viscosity of the Earth's inner core: Constraints from nutation observations. *Earth and Planetary Science Letters* 308: 343–349. http://dx.doi.org/10.1016/j.epsl.2011.06.004.
Koot L, Dumberry M, Rivoldini A, de Viron O, and Dehant V (2010) Constraints on the coupling at the core–mantle and inner core boundaries inferred from nutation observations. *Geophysical Journal International* 182: 1279–1294. http://dx.doi.org/10.1111/j.1365-246X.2010.04711.x.
Koot L, Rivoldini A, de Viron O, and Dehant V (2008) Estimation of Earth interior parameters from a Bayesian inversion of VLBI nutation time series. *Journal of Geophysical Research* 113(B8): B08414. http://dx.doi.org/10.1029/2007JB005409.
Lambert S and Mathews PM (2005) Second order torque on the tidal redistribution and the Earth's rotation. *Astronomy and Astrophysics* (submitted for publication).
Legros H, Hinderer J, Lefftz M, and Dehant V (1993) The influence of the solid inner core on gravity changes and spatial nutations induced by luni-solar tides and surface loading. *Physics of the Earth and Planetary Interiors* 76: 283–315.
Mathews PM and Bretagnon P (2003) *Astronomy and Astrophysics* 400: 1113–1128.
Mathews PM, Buffett BA, Herring TA, and Shapiro II (1991) Forced nutations of the Earth: Influence of inner core dynamics. I. Theory. *Journal of Geophysical Research* 96(B5): 8219–8242.
Mathews PM and Guo JY (2005) Visco-electromagnetic coupling in precession-nutation theory. *Journal of Geophysical Research* 110: B02402. http://dx.doi.org/10.1029/2003JB002915.
Mathews PM, Herring TA, and Buffett BA (2002) Modeling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior. *Journal of Geophysical Research* 107(B4). http://dx.doi.org/10.1029/2001JB000390.
McCarthy DD (1996) IERS Conventions 1996, IERS Technical Note 22.
McCarthy DD (2003) IERS Conventions 2000, IERS Technical Note 29.
Molodensky MS (1961) The theory of nutation and diurnal Earth tides. *Communications de l'Observatoire Royal de Belgique* 188: 25–66.
Mueller II (1981) Reference coordinate systems for earth dynamics: A preview. In: Gaposchkin EM and Kolaczek B (eds.) *Reference Coordinate Systems for Earth Dynamics*, pp. 1–22. Dordrecht, Netherlands: Reidel Publishing Company.
Poincaré H (1910) Sur la precession des corps deformables. *Bulletin Astronomique* 27: 321–356.
Rogister Y (2001) On the diurnal and nearly diurnal free modes of the Earth. *Geophysical Journal International* 144(2): 459–470.
Roosbeek F (1999) Diurnal and subdiurnal terms in Rdan97 series. *Celestial Mechanics and Dynamical Astronomy* 74(4): 243–252.
Roosbeek F and Dehant V (1998) RDAN97: An analytical development of rigid Earth nutation series using the torque approach. *Celestial Mechanics and Dynamical Astronomy* 70: 215–253.
Rutter MD, Secco RA, Uchida T, et al. (2002) Towards evaluating the viscosity of the Earth's outer core: An experimental high pressure study of liquid Fe–S (8.5 wt. per cent S). *Geophysical Research Letters* 29(8): 080000–080001.
Sasao T, Okubo S, and Saito M (1980) A simple theory on the dynamical effects of a stratified fluid core upon nutational motion of the Earth. In: *Nutation and the Earth's Rotation*. Proceedings of the IAU Symposium, Kiev, Ukrainian SSR, May 23–28, 1977, pp. 165–183. Dordrecht, Netherlands: D. Reidel Publishing Co.
Sasao T and Wahr JM (1981) An excitation mechanism for the free core nutation. *Geophysical Journal of the Royal Astronomical Society* 64: 729–746.
Smith M (1974) The scalar equations of infinitesimal elastic-gravitational motion for a rotating, slightly elliptical Earth. *Geophysical Journal of the Royal Astronomical Society* 37: 491–526.
Soffel M and Klioner S (1998) The present status of Einstein relativistic celestial mechanics. In: Vondrak J and Capitaine N (eds.) *Journées Systèmes de Références, September 1997, Prague, Czech Republic*, pp. 27–31.
Souchay J and Folgueira M (2000) The effect of zonal tides on the dynamical ellipticity of the Earth and its influence on the nutation. *Earth, Moon, and Planets* 81: 201–216.
Souchay J and Kinoshita H (1996) Corrections and new developments in rigid Earth nutation theory: I. Lunisolar influence including indirect planetary effects. *Astronomy and Astrophysics* 312: 1017–1030.
Souchay J and Kinoshita H (1997) Corrections and new developments in rigid Earth nutation theory: II. Influence of second-order geopotential and direct planetary effect. *Astronomy and Astrophysics* 318: 639–652.
Souchay J, Loysel B, Kinoshita H, and Folgueira M (1999) Corrections and new developments in rigid Earth nutation theory, III. Final tables REN2000 including crossed nutation and spin–orbit coupling effects. *Astronomy and Astrophysics Supplement Series* 135: 111–139.
Trinh A, Rivoldini A, Van Hoolst T, and Dehant V (2011) The librations of a triaxial, synchronously rotating planetary satellite. In: *EPSC-DPS Joint Meeting 2011, La Cité Internationale des Congrès Nantes Métropole, Nantes, France, 3–7 October 2011*, extended abstract, EPSC Proceedings, 6, EPSC-DPS2011-1514, 2 pages.
Wahr JM (1981) The forced nutations of an elliptical, rotating, elastic and oceanless Earth. *Geophysical Journal of the Royal Astronomical Society* 64: 705–727.
Wahr JM and Bergen Z (1986) The effects of mantle anelasticity on nutations, Earth tides and tidal variations in rotation rate. *Geophysical Journal of the Royal Astronomical Society* 87: 633–668.
Wahr JM and Sasao T (1981) A diurnal resonance in the ocean tide and in the Earth's load response due to the resonant free "core nutation" *Geophysical Journal of the Royal Astronomical Society* 64: 747–765.
Widmer R, Masters G, and Gilbert F (1991) Spherically symmetric attenuation within the Earth from normal mode data. *Geophysical Journal International* 104: 541–553.
Williams JG (1994) Contributions to the Earth's obliquity rate, precession, and nutation. *Astronomy Journal* 108: 711–724.
Wu X and Wahr JM (1997) Effects of non-hydrostatic core–mantle boundary topography and core dynamics on Earth rotation. *Geophysical Journal International* 128: 18–42.

This page intentionally left blank

3.11 GPS and Space-Based Geodetic Methods

G Blewitt, University of Nevada, Reno, NV, USA

© 2015 Elsevier B.V. All rights reserved.

3.11.1	**The Development of Space Geodetic Methods**	307
3.11.1.1	Introduction	307
3.11.1.2	The Limitations of Classical Surveying Methods	308
3.11.1.3	The Impact of Space Geodesy	309
3.11.1.4	Lunar Laser Ranging Development	310
3.11.1.5	Satellite Laser Ranging Development	310
3.11.1.6	VLBI Development	311
3.11.1.7	Global Positioning System Development	311
3.11.1.8	Comparing GPS with VLBI and SLR	313
3.11.1.9	GPS Receivers in Space: LEO GPS	314
3.11.1.10	Global Navigation Satellite Systems	314
3.11.1.11	International GNSS Service	315
3.11.2	**GPS and Basic Principles**	316
3.11.2.1	Basic Principles	316
3.11.2.2	GPS Design and Consequences	318
3.11.2.3	Introducing High-Precision GPS	319
3.11.2.4	GPS Observable Modeling	321
3.11.2.5	Data Processing Software	324
3.11.2.6	Real-Time GPS and Accuracy Versus Latency	326
3.11.3	**Global and Regional Measurement of Geophysical Processes**	326
3.11.3.1	Introduction	326
3.11.3.2	Estimation of Station Velocity	328
3.11.3.3	Plate Tectonic Rotations	329
3.11.3.4	Plate Boundary Strain Accumulation	330
3.11.3.5	The Earthquake Cycle	332
3.11.3.6	Surface Mass Loading	334
References		335

3.11.1 The Development of Space Geodetic Methods

3.11.1.1 Introduction

Geodesy is the science of accurately measuring and understanding three fundamental properties of the Earth: (1) its gravitational field, (2) its geometric shape, and (3) its orientation in space (Torge, 2001). In recent decades, the growing emphasis has been on the time variation of these 'three pillars of geodesy' (Beutler et al., 2004), which has become possible owing to the accuracy of new space-based geodetic methods and also owing to a truly global reference system that only space geodesy can realize (Altamimi et al., 2001, 2002). As each of these three properties is connected by physical laws and is forced by natural processes of scientific interest (Lambeck, 1988), thus, space geodesy has become a highly interdisciplinary field, intersecting with a vast array of geophysical disciplines, including tectonics, Earth structure, seismology, oceanography, hydrology, atmospheric physics, meteorology, and climate change. This richness of diversity has provided the impetus to develop space geodesy as a precise geophysical tool that can probe the Earth and its interacting spheres in ways never before possible (Smith and Turcotte, 1993).

Borrowing from the fields of navigation and radio astronomy and classical surveying, space geodetic methods were introduced in the early 1970s with the development of lunar laser ranging (LLR), satellite laser ranging (SLR), and very long baseline interferometry (VLBI) and soon to be followed by the Global Positioning System (GPS) (Smith and Turcotte, 1993). The near future promises other new space geodetic systems similar to GPS, which can be more generally called Global Navigation Satellite Systems (GNSSs). In recent years, the GPS has become commonplace, serving a diversity of applications from car navigation to surveying. Originally designed for few meter-level positioning for military purposes, GPS is now routinely used in many areas of geophysics (Bilham, 1991; Dixon, 1991; Hager et al., 1991; Segall and Davis, 1997), for example, to monitor the movement of the Earth's surface between points on different continents with millimeter-level precision, essentially making it possible to observe plate tectonics as it happens.

The stringent requirements of geophysics are part of the reason as to why GPS has become as precise as it is today (Blewitt, 1993). As will be described here, novel techniques have been developed by researchers working in the field of geophysics and geodesy, resulting in an improvement of GPS precision by four orders of magnitude over the original design specifications. Owing to this high level of precision and the relative ease of acquiring GPS data, GPS has revolutionized geophysics, as well as many other areas of human endeavor (Minster et al., 2010).

Whereas perhaps the general public may be more familiar with the georeferencing applications of GPS, say, to locate a vehicle on a map, this chapter introduces space geodetic methods with a special focus on GPS as a high-precision geodetic technique and introduces the basic principles of geophysical research applications that this capability enables. As an example of the exacting nature of modern GPS geodesy, **Figure 1** shows a geodetic GPS station of commonplace (but leading-edge) design now in the western United States, installed for purposes of measuring tectonic deformation across the boundary between the North American and Pacific plates. This station was installed in 1996 at Slide Mountain, Nevada, as part of the BARGEN network (Bennett et al., 1998, 2003; Wernicke et al., 2000). To mitigate the problem of very local, shallow surface motions (Wyatt, 1982), this station has a deep brace Wyatt-type monument design, by which the antenna is held fixed to the Earth's crust by four welded braces that are anchored ~10 m below the surface (and are decoupled by padded boreholes from the surface). Tests have shown that such monuments exhibit less environmentally caused displacement than those installed to a (previously more common) depth of ~2 m (Langbein et al., 1995). Time series of daily coordinate estimates from such sites indicate repeatability at the level of 1 mm horizontal, and 3 mm vertical, with a velocity uncertainty of 0.2 mm year^{-1} (Davis et al., 2003). This particular site detected ~10 mm of transient motion for 5 months in late 2003, concurrent with unusually deep seismicity below Lake Tahoe that was likely caused by intrusion of magma into the lower crust (Smith et al., 2004).

3.11.1.2 The Limitations of Classical Surveying Methods

It is useful to consider the historical context of terrestrial surveying at the dawn of modern space geodesy around 1970 (Bomford, 1971). Classical geodetic work of the highest (~mm) precision was demonstrated in the 1970s for purposes of measuring horizontal crustal strain over regional scales (e.g., Savage, 1983). However, the limitations of classical geodesy discussed in the succeeding text implied that it was essentially impossible to advance geodetic research on the global scale.

Figure 1 Permanent IGS station at Slide Mountain, Nevada, the United States. Photo by Jean Dixon.

Classical surveying methods were not truly three-dimensional. This is because geodetic networks were distinctly separated into horizontal networks and height networks, with poor connections between them. Horizontal networks relied on the measurement of angles (triangulation) and distances (trilateration) between physical points (or 'benchmarks') marked on top of triangulation pillars, and vertical networks mainly depended on spirit leveling between height benchmarks. In principle, height networks could be loosely tied to horizontal networks by collocation of measurement techniques at a subset of benchmarks, together with geometric observations of vertical angles. Practically, this was difficult to achieve, because of the differing requirements on the respective networks. Horizontal benchmarks could be separated further apart on hill tops and peaks, but height benchmarks were more efficiently surveyed along valleys wherever possible. Moreover, the measurement of angles is not particularly precise and is subject to significant systematic error, such as atmospheric refraction.

Fundamentally, however, the height measured with respect to the gravitational field (by spirit leveling) is not the same quantity as the geometric height, which is given relative to some conventional ellipsoid (that in some average sense represents sea level). Thus, the horizontal and height coordinate systems (often called a '2 + 1' system) could never be made entirely consistent.

A troublesome aspect of terrestrial surveying methods was that observations were necessarily made between benchmarks that were relatively close to each other, typically between nearest neighbors in a network. Because of this, terrestrial methods suffered from increase in errors as the distance increased across a network. Random errors would add together across a network, growing as a random walk process, proportional to the square root of distance.

Even worse, systematic errors in terrestrial methods (such as errors correlated with elevation, temperature, latitude, etc.) can grow approximately linearly with distance. For example, wave propagation for classical surveying occurs entirely within the troposphere, and thus, errors due to refraction increase with the distance between stations. In contrast, no matter how far apart the stations are, wave propagation for space geodetic techniques occurs almost entirely in the approximate vacuum of space and is only subject to refraction within ~10 km optical thickness of the troposphere (and within the ionosphere in the case of microwave techniques, although ionospheric refraction can be precisely calibrated by dual-frequency measurements). Furthermore, by modeling the changing slant depth through the troposphere (depending on the source position in the sky), tropospheric delay can be accurately estimated as part of the positioning solution.

There were other significant problems with terrestrial surveying that limited its application to geophysical geodesy. One was the requirement of interstation visibility, not only with respect to intervening terrain but also with respect to the weather at the time of observation. Furthermore, the precision and accuracy of terrestrial surveying depended a lot on the skill and experience of the surveyors making the measurements and the procedures developed to mitigate systematic error while in the field (i.e., errors that could not readily be corrected after the fact).

Finally, the spatial extent of classical terrestrial surveying was limited by the extent of continents. In practice, different

countries often adopted different conventions to define the coordinates of their national networks. As a consequence, each nation typically had a different reference system. More importantly from a scientific viewpoint, connecting continental networks across the ocean was not feasible without the use of satellites. So in the classical geodetic era, it was possible to characterize the approximate shape of the Earth; however, the study of the change of the Earth's shape in time was for all practical purposes out of the question.

3.11.1.3 The Impact of Space Geodesy

Space geodetic techniques have since solved all the aforementioned problems of terrestrial surveying. Therefore, the impact of space geodetic techniques can be summarized as follows (as will be explained in detail later):

- They allow for true three-dimensional positioning.
- They allow for relative positioning that does not degrade significantly with distance.
- They do not require interstation visibility and can tolerate a broader range of weather conditions.
- The precision and accuracy are far superior and position estimates are more reproducible and repeatable than for terrestrial surveying, where for space geodesy, the quality is determined more by the quality of the instruments and data processing software than by the skill of the operator.
- They allow for global networks that can define a global reference frame; thus, the position coordinates of stations in different continents can be expressed in the same system.

From a geophysical point of view, the advantages of space geodetic techniques can be summarized as follows:

- The high precision of space geodesy (now at the $\sim$1 mm level), particularly over very long distances, allows for the study of Earth processes that could never be observed with classical techniques.
- The Earth's surface can be surveyed in one consistent reference frame, so geophysical processes can be studied in a consistent way over distance scales ranging ten orders of magnitude from 10° to 10^{10} m (Altamimi et al., 2002). Global surveying allows for the determination of the largest-scale processes on Earth, such as plate tectonics and surface mass loading.
- Geophysical processes can be studied in a consistent way over timescales ranging ten orders of magnitude from 10^{-1} to 10^{9} s. Space geodetic methods allow for continuous acquisition of data using permanent stations with communications to a data processing center. This allows for geophysical processes to be monitored continuously, which is especially important for the monitoring of natural hazards but is also important for the characterization of errors and for the enhancement of precision in the determination of motion. Sample rates from GPS can be as high as 50 Hz. Motion is fundamentally determined by space geodesy as a time series of positions relative to a global reference frame. Precise timing of the sampled positions in a global timescale (<0.1 μs UTC in even the most basic form of GPS and <0.1 ns for geodetic GPS) is an added bonus for some applications, such as seismology and SLR.
- Space geodetic surveys are more cost-efficient than classical methods; thus, more points can be surveyed over a larger area than was previously possible.

The benefits that space geodesy could bring to geophysics are precisely the reason why space geodetic methods were developed. For example, NASA's interest in directly observing the extremely slow motions (centimeters per year) caused by plate tectonics was an important driver in the development of SLR, geodetic VLBI, and geodetic GPS (Smith and Turcotte, 1993). SLR was initially a NASA mission dedicated to geodesy. VLBI and GPS were originally developed for other purposes (astronomy and navigation, respectively), though with some research and development (motivated by the potential geophysical reward) they were adapted into high-precision geodetic techniques for geophysical research.

The following are just a few examples of geophysical applications of space geodesy:

- Plate tectonics, by tracking the relative rotations of clusters of space geodetic stations on different plates.
- Interseismic strain accumulation, by tracking the relative velocity between networks of stations in and around plate boundaries.
- Earthquake rupture parameters, by inverting measurements of coseismic displacements of stations located within a few rupture lengths of the fault.
- Postseismic processes and rheology of the Earth's topmost layers, by inverting the decay signature (exponential, logarithmic, etc.) of station positions in the days to decades following an earthquake.
- Magmatic processes, by measuring time variation in the position of stations located on volcanoes or other regions of magmatic activity, such as hot spots.
- Rheology of the Earth's mantle and ice-sheet history, by measuring the vertical and horizontal velocities of stations in the area of postglacial rebound (glacial isostatic adjustment).
- Mass redistribution within the Earth's fluid envelope, by measuring time variation in the Earth's shape, the velocity of the solid Earth's center of mass, the Earth's gravitational field, and the Earth's rotation in space.
- Global change in sea level, by measuring vertical movement of the solid Earth at tide gauges, by measuring the position of spaceborne altimeters in a global reference frame, and by inferring exchange of water between the oceans and continents from mass redistribution monitoring.
- Hydrology of aquifers by monitoring aquifer deformation inferred from time variation in 3-D coordinates of a network of stations on the surface above the aquifer.
- Providing a global reference frame for consistent georeferencing and precision time tagging of nongeodetic measurements and sampling of the Earth, with applications in seismology, airborne and spaceborne sensors, and general fieldwork.

What characterizes modern space geodesy is the broadness of its application to almost all branches of geophysics and the pervasiveness of geodetic instrumentation and data used by geophysicists who are not necessarily experts in geodesy. GPS provides easy access to the global reference frame, which in turn fundamentally depends on the complementary benefits of

all space geodetic techniques (Herring and Perlman, 1993). In this way, GPS provides access to the stability and accuracy inherent in SLR and VLBI without the need for coordination on the part of the field scientist. Moreover, GPS geodesy has benefited tremendously from earlier developments in SLR and VLBI, particularly in terms of modeling the observations.

3.11.1.4 Lunar Laser Ranging Development

Geodesy was launched into the space age by LLR, a pivotal experiment in the history of geodesy. The basic concept of LLR is to measure the distance to the Moon from an Earth-based telescope by timing the flight of a laser pulse that is emitted by the telescope, reflects off the Moon's surface, and is received back into the same telescope. LLR was enabled by the Apollo 11 mission in July 1969, when Buzz Aldrin deployed a laser retroreflector array on the Moon's surface in the Sea of Tranquility (Dickey et al., 1994). Later, Apollo 14 and 15 and a Soviet Lunokhod mission carrying French-built retroreflectors have expanded the number of sites on the Moon. Since initial deployment, several LLR observatories have recorded measurements around the globe, although most of the routine observations have been made at only two observatories: McDonald Observatory in Texas, the United States, and the CERGA station in France. Today, the McDonald Observatory uses a 0.726 m telescope with a frequency-doubled neodymium-doped–YAG laser, producing 1500 mJ pulses of 200 ps width at 532 nm wavelength, at a rate of 10 Hz.

The retroreflectors on the lunar surface are corner cubes, which have the desirable property that they reflect light in precisely the opposite direction, independent of the angle of incidence. Laser pulses take between 2.3 and 2.6 s to complete the 385 000 km journey. The laser beam width expands from 7 mm on Earth to several km at the Moon's surface (a few km), and so in the best conditions, only one photon of light will return to the telescope every few seconds. By timing the flight of these single photons, ranges to the Moon can now be measured with a precision approaching 1 cm.

The LLR experiment has produced the following important research findings fundamental to geophysics (Williams et al., 2001, 2004), all of which represent the most stringent tests to date:

- The Moon is moving radially away from the Earth at 38 mm year^{-1}, an effect attributed to tidal friction, which slows down the Earth's rotation, hence increasing the Moon's distance so as to conserve angular momentum of the Earth–Moon system.
- The moon likely has a liquid core.
- The Newtonian gravitational constant G is stable to $< 10^{-12}$.
- Einstein's theory of general relativity correctly explains the Moon's orbit to within the accuracy of LLR measurements. For example, the equivalence principle is verified with a relative accuracy of 10^{-13}, and geodetic precession is verified to within $< 0.2\%$ of general relativistic expectations.

3.11.1.5 Satellite Laser Ranging Development

SLR was developed in parallel with LLR and is based on similar principles, with the exception that the retroreflectors (corner cubes) are placed on artificial satellites (Degnan, 1993). Experiments with SLR began in 1964 with NASA's launch of the Beacon-B satellite, tracked by Goddard Space Flight Center with a range accuracy of several meters. Following a succession of demonstration tests, operational SLR was introduced in 1975 with the launch of the first dedicated SLR satellite, Starlette, launched by the French Space Agency, soon followed in 1976 by NASA's Laser Geodynamics Satellite (LAGEOS-1) in a near-circular orbit of 6000 km radius. Since then, other SLR satellites now include LAGEOS-II, Stella, Etalon-1 and -2, and AJISAI. There are now approximately ten dedicated satellites that can be used as operational SLR targets for a global network of more than 40 stations, most of them funded by NASA for purposes of investigating geodynamics, geodesy, and orbital dynamics (Tapley et al., 1993).

SLR satellites are basically very dense reflecting spheres orbiting the Earth. For example, LAGEOS-II launched in 1992 is a 0.6 m sphere of mass 411 kg. The basic principle of SLR is to time the round-trip flight of a laser pulse shot from the Earth to the satellite. Precise time tagging of the measurement is accomplished with the assistance of GPS. The round-trip time of flight measurements can be made with centimeter-level precision, allowing for the simultaneous estimation of the satellite orbits, gravitational field parameters, tracking station coordinates, and Earth rotation parameters. The reason the satellites have been designed with a high mass to surface area ratio is to minimize accelerations due to nonconservative forces such as drag and solar radiation pressure. This produces a highly stable and predictable orbit and hence a stable dynamic frame from which to observe the Earth's rotation and station motions.

SLR made early contributions to the confirmation of the theory of plate tectonics (Smith et al., 1993) and toward measuring and understanding contemporary crustal deformation in plate boundary zones (Jackson et al., 1994; Wilson and Reinhart, 1993). To date, SLR remains the premier technique for determining the location of the center of mass of the Earth system and its motion with respect to the Earth's surface (Chen et al., 1999; Ray, 1998; Watkins and Eanes, 1997). As an optical technique that is relatively less sensitive to water vapor in the atmosphere, SLR has also played a key role in the realization of reference frame scale (Dunn et al., 1999). The empirical realization of scale and origin is very important for the testing of dynamic Earth models within the rigorous framework of the International Terrestrial Reference System (ITRS) (McCarthy, 1996).

Today, SLR is used in the following research (Perlman et al., 2002):

- Mass redistribution in the Earth's fluid envelope, allowing for the study of atmosphere–hydrosphere–cryosphere–solid Earth interactions. SLR can sense the Earth's changing gravitational field (Bianco et al., 1997; Cheng and Tapley, 1999, 2004; Gegout and Cazenave, 1993; Nerem et al., 1993; Devoti et al., 2001), the location of the solid Earth's center of mass with respect to the center of mass of the entire Earth system (Chen et al., 1999). Also, SLR determination of the Earth's rotation in the frame of the stable satellite orbits reveals the exchange of angular momentum between the solid Earth and fluid components of the Earth

system (Chao et al., 1987). SLR stations can sense the deformation of the Earth's surface in response to loading of the oceans, atmosphere, and hydrosphere and can infer mantle dynamics from response to the unloading of ice from past ice ages (Argus et al., 1999).

- Long-term dynamics of the solid Earth, oceans, and ice fields (Sabadini et al., 2002). SLR can sense surface elevations unambiguously with respect to the Earth's center of mass, such as altimeter satellite height and hence ice-sheet and sea surface height. Thus, SLR is fundamental to the terrestrial reference frame and the long-term monitoring of sea level change.
- Mantle–core interaction through long-term variation in the Earth's rotation (Eubanks, 1993).
- General relativity, specifically the Lense–Thirring effect of frame dragging (Cuifolini and Pavlis, 2004).

SLR is a relatively expensive and cumbersome technique and so has largely been superseded by the GPS technique for most geophysical applications. SLR is still necessary for maintaining the stability of the International Terrestrial Reference Frame (ITRF), in particular, to aligning the ITRF origin with the specifications of ITRS (Altamimi et al., 2002). SLR is also necessary to determine long-term variation in the low-degree components of the Earth's gravitational field. SLR is maintained by NASA to support high-precision orbit determination (such as for satellite altimetry), though GPS is also now being used for that purpose.

3.11.1.6 VLBI Development

VLBI, originally a technique designed for observing distant celestial radio sources with high angular resolution, was from the late 1970s developed for high-precision geodetic applications by applying the technique 'in reverse' (Rogers et al., 1978). Much of this development of geodetic VLBI was performed by the NASA Crustal Dynamics Project initiated in 1979 (Bosworth et al., 1993) with the idea to have an alternative technique to SLR to provide independent confirmation of scientific findings.

Conceptually, geodetic VLBI uses radio waves from distant quasars at known positions on the celestial sphere and measures the difference in the time of arrival of signals from those quasars at stations (radio observatories) on the Earth's surface. Such data provide information on how the geometry of a network of stations evolves in time. This time-variable geometry can be inverted to study geophysical processes such as the Earth's rotation and plate tectonics and can be used to define a global terrestrial reference frame with high precision. Unique to VLBI is that it can provide an unambiguous, stable tie between the orientation of the terrestrial reference frame and the celestial reference frame, that is, Earth orientation. However, as a purely geometric technique, it is not directly sensitive to the Earth's center of mass and gravitational field, although inferences by VLBI on gravity can be made through models that connect gravity to Earth's shape, such as tidal and loading models.

Comparisons between VLBI and SLR proved to be important for making improvements in both methods. As a radio technique, VLBI is more sensitive to errors in atmospheric refraction (Davis et al., 1985; Niell, 1996; Truehaft and Lanyi, 1987) than the optical SLR technique; however, VLBI has the advantage that the sources are quasars that appear to be essentially fixed in the sky, thus providing the ultimate in celestial reference frame stability. VLBI is therefore the premier technique for determining parameters describing the Earth rotation's in inertial space, namely, precession, nutation, and UT1 (the angle of rotation with respect to UTC) (Eubanks, 1993). VLBI ultimately has proven to be more precise than SLR in measuring distances between stations.

However, VLBI has never been adapted for tracking Earth-orbiting platforms and is highly insensitive to the Earth's gravitational field and thus cannot independently realize the Earth's center of mass as the origin of the global reference frame. On the other hand, the stability of scale in VLBI is unsurpassed. For most geophysical applications, GPS has superseded VLBI, except for the important reference frame and Earth orientation tasks described earlier. VLBI remains important for characterizing long-wavelength phenomena such as postglacial rebound, with the highest precision among all techniques today, and therefore is integral to the stability of global terrestrial reference frames.

To summarize, geodetic VLBI's main contributions to scientific research involve (Schlueter et al., 2002)

- unambiguous Earth orientation parameters, which can be used to study angular momentum exchange between the solid Earth and its fluid reservoirs and provides a service to astronomy and space missions by connecting the terrestrial reference frame to the celestial reference frame (Eubanks, 1993);
- providing a stable scale for the global terrestrial reference frame (Boucher and Altamimi, 1993);
- providing the highest-precision measurements of long-wavelength Earth deformations, thus providing stability to the global frame, and constraints on large-scale geodynamics such as postglacial rebound and plate tectonics (Argus et al., 1999; Stein, 1993).

3.11.1.7 Global Positioning System Development

As of September 2006, the GPS consists of 29 active satellites that can be used to position a geodetic receiver with an accuracy of millimeters within the ITRF. To do this requires geodetic-class receivers (operating at two frequencies, and with antennas designed to suppress signal multipath), currently costing a few thousand US dollars, and geodetic research-class software (developed by various universities and government institutions around the world). Such software embody leading-edge models (of the solid Earth, atmosphere, and satellite dynamics) and data processing algorithms (signal processing and stochastic parameter estimation). Many of the models have been developed as a result of much research conducted by the international geodetic and geophysical community, often specifically to improve the accuracy of GPS. Today, it is even possible for a nonexpert to collect GPS data and produce receiver positions with centimeter accuracy by using an Internet service for automatic data processing.

The geodetic development of the GPS has been driven by a number of related factors (Blewitt, 1993):

- The foundation for many of the research-class models was already in place owing to the similarities between GPS and VLBI (as radio techniques) and GPS and SLR (as satellite dynamic techniques), thus giving an early boost to GPS geodesy. Continued collaboration with the space geodetic community has resulted in standard models such as those embodied by the ITRS Conventions (McCarthy, 1996), which aim to improve the accuracy and compatibility of results from the various space geodetic techniques.
- GPS is of relatively low cost and yet has comparable precision to VLBI and SLR. Whereas the GPS itself is paid for by the US taxpayer, the use of the system is free to all as a public good. This has made GPS accessible to university researchers, and the resulting research has further improved GPS accuracy through better models.
- GPS stations are easy to deploy and provide a practical way to sample the deformation field of the Earth's surface more densely, thus allowing space geodesy to address broader diversity scientific questions. This has opened up interdisciplinary research within geophysics, leading to discoveries in unforeseen areas and to further improvements in GPS accuracy through improved observation models.
- GPS was readily adopted because of the ease of access to the ITRF on an ad hoc basis, without the need for special global coordination from the point of view of an individual investigator. Furthermore, the ITRF gives implicit access to the best possible accuracy and stability that can be achieved by SLR and VLBI (Herring and Pearlman, 1993).

Following closely the historical perspectives of Evans et al. (2002) and Blewitt (1993), GPS has its roots as a successor to military satellite positioning systems developed in the 1960s, though the first geophysical applications of GPS were not realized until the early 1980s. In the run-up to the space age in 1955, scientists at the Naval Research Laboratory first proposed the application of satellite observations to geodesy. By optical observation methods, the first geodetic satellites were quickly used to refine parameters of the Earth's gravitational field. Optical methods were eventually made obsolete by the Doppler technique employed by the Navy Navigation Satellite System (TRANSIT). As the name implies, Doppler positioning was based on measuring the frequency of the satellite signal as the relative velocity changed between the satellite and the observer. By the early 1970s, Doppler positioning with 10 m accuracy became possible on the global scale, leading to the precise global reference frame 'World Geodetic System 1972' (WGS 72), further improved by WGS 84, which was internally accurate at the 10 cm level. Having a global network of known coordinates together with the success of radiometric tracking methods set the stage for the development of a prototype GPS in the late 1970s.

The US Department of Defense launched its first prototype Block I GPS satellite, NAVSTAR 1, in February 1978. By 1985, ten more Block 1 satellites had been launched, allowing for the development and testing of prototype geodetic GPS data processing software that used dual-frequency carrier phase observables. In February 1989, the first full-scale operational GPS satellite known as Block II was deployed, and by January 1994, a nominally full constellation of 24 satellites was completed, ensuring that users could see satellites of a sufficient number (at least 5) at anytime, anywhere in the world. Initial operational capability was officially declared in December 1993, and full operational capability was declared in April 1995. From July 1997, Block IIRs began to replace GPS satellites. The first modified version Block IIR-M satellite was launched in 2005 (for the first time emitting the L2C signal, which allows civilian users to calibrate for ionospheric delay). The current constellation of 29 satellites includes extra satellites as 'active spares' to ensure seamless and rapid recovery from a satellite failure. The first Block IIF satellite was scheduled to launch in 2008 and may transmit a new civil signal at a third frequency.

The GPS design built on the success of Doppler by enabling the measurement of a biased range ('pseudorange') to the satellite, which considerably improved positioning precision. Carrier phase tracking technology further improved the signal measurement precision to the few millimeter level. As a radio technique, VLBI technology was adapted in NASA's prototype geodetic GPS receivers. The SERIES receiver, developed by MacDoran (1979) at the Jet Propulsion Laboratory (JPL), pointed at one source at a time using a directional antenna (a technique no longer used). Many key principles and benefits of the modern GPS geodesy were based on the omnidirectional instrument, MITES, proposed by Counselman and Shapiro (1979). This was developed by the Massachusetts Institute of Technology (MIT) group into the Macrometer instrument, which provided centimeter-level accuracy using the innovative double-difference method for eliminating clock bias, a method that has its origins in radio navigation of the Apollo mission (Counselman et al., 1972).

By the mid 1980s, commercial receivers such as the Texas Instrument TI 4100 became available (Henson et al., 1985) and were quickly deployed by geophysicists in several pioneering experiments to measure the slow motions associated with plate tectonics (Dixon et al., 1985; Freymueller and Kellogg, 1990; Prescott et al., 1989); such experiments spurred the development of analysis techniques to improve precision at the level required by geophysics (Larson and Agnew, 1991; Larson et al., 1991; Tralli et al., 1988). Important developments during these early years include ambiguity resolution over long distances (Blewitt, 1989; Dong and Bock, 1989), precise orbit determination (Beutler et al., 1985; King et al., 1984; Lichten and Border, 1987; Swift, 1985), and troposphere modeling (Davis et al., 1987; Lichten and Border, 1987; Tralli and Lichten, 1990).

The development of geodetic GPS during the 1980s was characterized by intensive hardware and software development with the goal of sub-centimeter positioning accuracy, over increasingly long distances. A prototype digital receiver known as 'Rogue' was developed by the JPL (Thomas, 1988), which produced high-precision pseudorange data that could be used to enhance data processing algorithms, such as ambiguity resolution. Several high-precision geodetic software packages that were developed around this time are still in use and far exceed the capabilities of commercial packages. These included the Bernese developed at the University of Bern (Beutler et al., 1985; Gurtner, 1985; Rothacher et al., 1990), GAMIT-GLOBK developed at MIT (Bock et al., 1986; Dong and Bock, 1989; Herring et al., 1990), and GIPSY-OASIS developed at JPL (Blewitt, 1989, 1990; Lichten and Border, 1987; Sovers and Border, 1987).

GPS became fully operational in 1994, with the completion of a full constellation of 24 satellites. Developments toward high precision in the 1990s include (1) truly global GPS solutions made possible by the completion of the Block II GPS constellation and, simultaneously, installation and operation of the global network in 1994 (shown in its current configuration in **Figure 2**) by the International GPS Service (IGS, since renamed the International GNSS Service) (Beutler et al., 1994a); (2) global-scale ambiguity resolution (Blewitt and Lichten, 1992); (3) further refinement to tropospheric modeling and the inclusion of tropospheric gradient parameters (Bar-Sever et al., 1998; Chen and Herring, 1997; Davis et al., 1993; McMillan, 1995; Niell, 1996; Rothacher et al., 1998); (4) adoption of baseband-digital GPS receivers with the low-multipath choke ring antenna developed originally at JPL (Meehan et al., 1992), which remains the IGS standard design today; (5) improved orbit models, particularly with regard to GPS satellite attitude, and the tuning of stochastic models for solar radiation pressure (Bar Sever, 1996; Beutler et al., 1994b; Fliegel and Gallini, 1996; Fliegel et al., 1992; Kuang et al., 1996); (6) improved reference system conventions (McCarthy, 1996); and (7) simultaneous solution for both orbits and station positions (fiducial-free global analysis) (Heflin et al., 1992).

The focus of developments in the decade includes (1) building on earlier work by Schupler et al. (1994), antenna phase center variation modeling and calibrations for both stations and the GPS satellites themselves (Ge et al., 2005; Mader, 1999; Mader and Czopek, 2002; Schmid and Rothacher, 2003; Schmid et al., 2005); (2) densification of stations in the ITRF and the installation of huge regional networks of geodetic GPS stations, such as the ~1000 station Plate Boundary Observatory currently being installed in the western North America (Silver et al., 1999); (3) improved analysis of large regional networks of stations through common-mode signal analysis (Wdowinski et al., 1997) and faster data processing algorithms ((Bertiger et al., 2010; Zumberge et al., 1997); (4) the move toward real-time geodetic analysis with applications such as GPS seismology (Larson et al., 2003; Nikolaidis et al., 2001) and tsunami warning systems (Blewitt et al., 2006), including signal processing algorithms to filter out sidereally repeating multipath (Bock et al., 2000, 2004; Choi et al., 2004); and (5) further improvements in orbit determination (Ziebart et al., 2002), tropospheric modeling (Boehm et al., 2006), and higher-order ionospheric models (Kedar et al., 2003).

With the significant improvements to modeling since the inception of the IGS in 1994, data reprocessing of global GPS data sets has begun in earnest. Early results indicate superior quality of the GPS data products, such as station coordinate time series, orbit and clock accuracy, and Earth orientation parameters (Steinberger et al., 2006).

3.11.1.8 Comparing GPS with VLBI and SLR

GPS geodesy can be considered a blend of the two earlier space geodetic techniques: VLBI and SLR. The most obvious similarities are that (1) SLR and GPS are satellite systems and so are sensitive to Earth's gravitational field and (2) VLBI and GPS are radio techniques and so the observables are subject to atmospheric refraction in a similar way. Due to these similarities, GPS geodesy has benefited from earlier work on both VLBI and SLR observation modeling and from reference system conventions already established through a combination of astronomical observation, VLBI and SLR observation, and geodynamic modeling. Moreover, Earth models such as tidal deformation are required by all global geodetic techniques, so GPS geodesy was in a position to exploit what had already been learned. Today, the improvements in modeling any of the techniques can often be exploited by another technique.

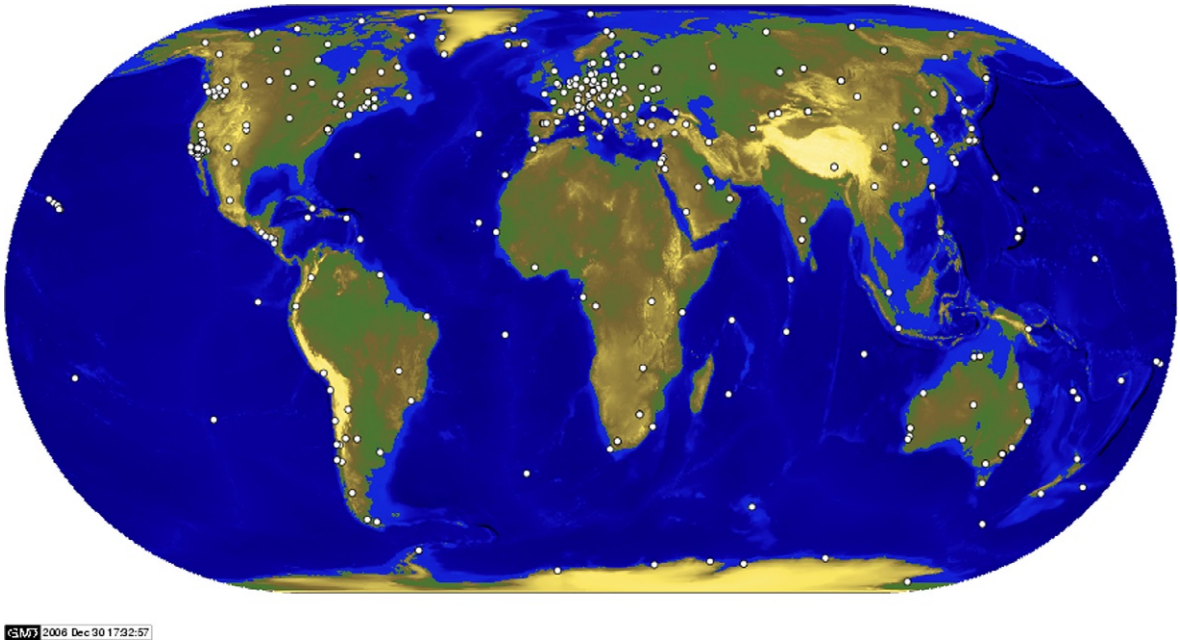


Figure 2 The global network of the International GPS Service. Courtesy of A. Moore.

The similarities between certain aspects of the different techniques lead to an overlap of strengths and weaknesses, error sources, and sensitivity to geophysical parameters of interest. The strengths and weaknesses can be summarized as follows:

- The main strength of SLR is the stability of the orbits due to custom-designed satellites. This leads to high sensitivity to the low-degree gravitational field harmonics and their long-term changes in time. This includes the degree-one term (characterizing geocenter motion, the motion of the solid Earth with respect to the center of mass of the entire Earth system), which is important for realizing the origin of the ITRF. As an optical technique, SLR is insensitive to moisture in the atmosphere and so has relatively small systematic errors associated with signal propagation. This inherently leads to more robust estimation of station height. On the other hand, SLR has problems working during daylight hours and in cloudy conditions. It is also an expensive and bulky technique and so suffers from a lack of geographic coverage.
- The main strength of VLBI is the stability and permanency of the sources, which are quasars. This leads to two important qualities: (1) VLBI is insensitive to systematic error in orbit dynamic models and can potentially be the most stable system for detecting the changes over the longest observed time periods, and (2) VLBI is strongly connected to an external, celestial reference frame, a vantage point from which Earth orientation and rotation can be properly determined. A major weakness of VLBI is (similar to SLR) its expensiveness and bulkiness. Moreover, some VLBI observatories are used for astronomical purposes and so cannot be dedicated to continuous geodetic measurement. VLBI antennas are very large structures that have their own set of problems, including the challenge to relate the observations to a unique reference point, and the stability of the structure with respect to wind and gravitational stress, and aging.

The main advantage of GPS is its low cost and ease of deployment and all-weather capability. Thus, GPS can provide much better geographic coverage, continuously. The flexibility of deployment allows for ties to be made between the terrestrial reference frames of the various techniques through collocation at SLR and VLBI sites. The disadvantage of GPS is that it is subject to both the systematic error associated with orbit dynamics and atmospheric moisture. Furthermore, the omnidirectional antennas of GPS lead to multipath errors. Thus, geodetic GPS is essential for improved sampling of the Earth in time and space but ultimately depends on SLR and VLBI to put such measurements into a reference frame that has long-term stability. This synergy lies at the heart of the emerging concept GGOS, the Global Geodetic Observing System, under the auspices of the International Association of Geodesy (IAG) (Rummel et al., 2005).

3.11.1.9 GPS Receivers in Space: LEO GPS

GPS has proved extremely important for positioning spaceborne scientific instruments in low-Earth orbit (sometimes called LEO GPS). Evans et al. (2002) provided an overview of spaceborne GPS, which is only briefly summarized here. The Landsat 4 satellite launched in 1982 was the first to carry a GPS receiver, called GPSPAC. This was followed by three more missions using GPSPAC, including Landsat 5 in 1984 and DoD satellites in 1983 and 1984. As the GPS satellite constellation grew during the 1980s, so the precision improved, enabling decimeter-level accuracy for positioning spaceborne platforms. Following Evans et al. (2002), the applications of spaceborne GPS can be categorized as (1) precise orbit determination of the host satellite for applications such as altimetry; (2) measurement of the Earth's gravitational field, such as the missions CHAMP and GRACE; (3) ionospheric imaging; and (4) indirect enhancements to global geodesy and remote sensing. In addition to these categories, spaceborne GPS is also being used to invert for the refractivity of the Earth's neutral atmosphere by occultation measurements, which can be used, for example, to infer stratospheric temperatures for studies of global climate change.

3.11.1.10 Global Navigation Satellite Systems

The success of GPS has led to the development of similar future systems, generically referred to as GNSS. To achieve global coverage, each GNSS system generally has a constellation of 20–30 satellites in roughly 12 h orbits. Some systems are augmented with a few satellites in either geostationary or inclined geosynchronous orbit.

The Russian system GLONASS (a Russian acronym that literally translates to GNSS) was actually developed in parallel with GPS and achieved global coverage with 24 satellites in orbit by 1995. After a subsequent period of degradation, the GLONASS system was restored back to a full constellation of 24 satellites by the end of 2011 and, as of 2013, has 29 satellites in orbit. Many modern GNSS receivers can track both GPS and GLONASS. Like GPS, the GLONASS satellite orbits and clocks are modeled by the IGS. However, in part due to the different transmission frequencies of the GLONASS satellites, which hinder the application of carrier phase ambiguity resolution techniques, the system has not proven to deliver geodetic solutions with such high precision as GPS. Nevertheless, GLONASS data can enhance GPS in situations where the sky is not completely visible, such as in an urban canyon environment.

An example of a GNSS under development is the European Galileo system, which has been scheduled to have full operational capability with 30 satellites before 2020, following several years of initial operational capability. By October 2012, four Galileo satellites had become operational, enabling the first production of 3-D position solutions.

The Chinese experimental regional BeiDou Navigation Satellite System (BDS) of five geostationary satellites is being expanded to have global coverage. BDS plans to add 30 non-geostationary satellites to the constellation, including three that are in inclined geosynchronous orbit. By 2013, BDS had 15 operational satellites and is planned to have a full global constellation by 2020.

Regional enhancement systems are being developed too. In Japan, the Quasi-Zenith Satellite System (QZSS) is planned to have three satellites in inclined geosynchronous orbit to enhance GPS in that region. As of 2012, one QZSS satellite

was in operation. Similarly, the Indian Regional Navigation Satellite System (IRNSS) is planned to have seven satellites to augment GPS (with three in geostationary orbit and four in inclined geosynchronous orbit), with the first launch planned for summer 2013.

The main reason for the development of alternative systems to GPS is to ensure access to GNSS signals that are not under the control of any single nation, with implications for the military in times of war and national emergencies and for civilian institutions such as national aviation authorities that have stringent requirements on guaranteed access to a sufficient number of GNSS signals at all times.

Thus, the future of GNSS is essentially guaranteed. By analogy with the Internet, navigation and geospatial referencing has become such an embedded part of the world's infrastructure and economy that it is now difficult to imagine a future world where GNSS is not pervasive. As GPS has proved, a GNSS system does not necessarily have to be designed with high-precision geodesy in mind in order for it to be used successfully as a high-precision geophysical tool. However, it is likely that future GNSS systems will take more into account the high-precision applications in their design and thus may be even better suited to geophysical applications than GPS currently is. Much can be done to mitigate errors, for example, in the calibration of the phase center variation in the satellite transmitting antenna or the transmission of signals at several different frequencies.

Satellite geodesy in the future will therefore use multiple GNSS systems interoperably and simultaneously. This will lead to improved precision and robustness of solutions. It will also allow for new ways to probe and hopefully mitigate systematic errors associated with specific GNSS systems and satellites. The continued downward spiral in costs of GNSS receiver systems will undoubtedly result in the deployment of networks with much higher density (reduced station spacing), which will benefit geophysical studies. For example, it would allow for higher-resolution determination of strain accumulation due to crustal deformation in plate boundary zones.

3.11.1.11 International GNSS Service

Infrastructure development and tremendous international cooperation characterized the 1990s. GPS operations moved away from the campaigns, back to the model of permanent stations, familiar to VLBI and SLR. As the prototype receivers developed by research groups in the 1980s had become commercialized, the cost of installing a GPS station in the 1990s had fallen to ~$25 K, in contrast to the millions of dollars required for VLBI/SLR. Thus, the long-range goal of the federal funding agencies was realized: dozens of GPS stations could be installed for the price of one VLBI station.

With the cooperation of ~100 research institutions around the world under the umbrella of the International GPS (now GNSS) Service (IGS), a global GPS network (now at ~350 stations; **Figure 2**) with a full geodetic analysis system came into full operation in 1994 (Beutler et al., 1994a). This backbone, together with the regional stations located in areas of tectonic activity, such as Japan and California, forms a global-scale instrument capable of resolving global plate tectonic motions and regional phenomena such as earthquake displacement. As a result of this international cooperation, a culture of data sharing has developed, with data freely available for research purposes via the Internet from IGS Global Data Centers. The establishment of a standard GPS measurement format known as RINEX (Receiver Independent Exchange) has facilitated this extensive exchange of data through IGS (see **Table 1** for IGS data availability).

The mission of the IGS is to provide the highest-quality data and products as the standard for GNSS in support of Earth science research and multidisciplinary applications. So although the IGS does not specifically carry out geophysical investigations, it does provide an essential service without which such investigations would be very costly and difficult to carry out. The 1990s has seen the development of collaborations with specific geophysical goals. Groups such as WEGENER (Plag et al., 1998) and UNAVCO (http://www.unavco.org) have provided an umbrella for geoscientists using GPS geodesy as a tool. Such groups depend on IGS for their success; conversely, IGS as a volunteer organization depends on such users to contribute to its operations and technical working groups.

The infrastructure has indeed become quite complex, yet cooperative, and often with an efficient division between geodetic operations and geodynamic investigations. As an example of how infrastructure is developing, solutions are being exchanged in a standard Software INdependent EXchange (SINEX) format to enable the construction of combined network solutions and, therefore, combined global solutions for Earth surface kinematics. This standard has since also been adopted by the other space geodetic techniques. Combination solutions have the advantage that (1) the processing burden is distributed among many groups who can check each other's solutions, (2) noise and errors are reduced through increased redundancy and quality control procedures, (3) coverage and density are increased, and (4) regional geodynamics can be interpreted in a self-consistent global context. An emerging focus of this decade (2000s) is the development of such combination solutions, and on the inversion of these solutions to infer geophysical parameters.

As the premier service for high-precision geodesy, the quality of IGS products is continually improving with time (**Figure 3**) and represents the current state-of-the-art (Dow

Table 1 IGS raw data types and availability

	Latency	*Updates*	*Sample interval*
Ground observations			
GPS and GLONASS data	1 day	Daily	30 s
	1 h	Hourly	30 s
	15 min	15 min	1 s[a]
GPS broadcast ephemerides	1 day	Daily	N/A
	1 h	Hourly	
	15 min	15 min	
GLONASS broadcast ephemerides	1 day	Daily	N/A
Meteorologic	1 day	Daily	5 min
	1 h	Hourly	5 min
Low-Earth orbiter observations			
GPS	4 days	Daily	10 s

[a]Selected subhourly stations have sampling intervals $1\text{ s} < t < 10\text{ s}$.
Source: IGS Central Bureau, http://igscb.jpl.nasa.gov.

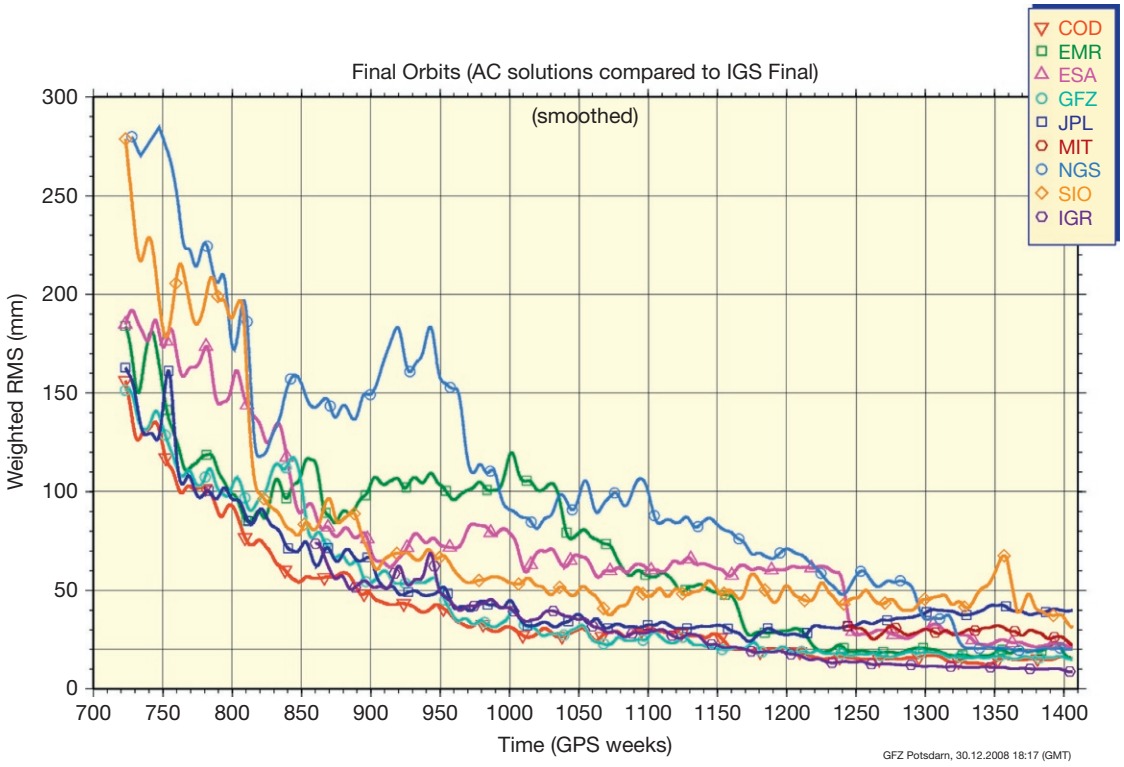


Figure 3 Plot showing the improvement of IGS orbit quality with time. Courtesy of G. Gendt.

et al., 2005b; Moore, 2007). The levels of accuracy claimed by the IGS for its various products are reproduced in Table 2.

Analogous to the IGS, geodetic techniques are organized as scientific services within the IAG. The IAG services are as follows:

- International Earth Rotation and Reference Systems Service (IERS) (IERS, 2004).
- International GNSS Service, formerly the International GPS Service (IGS) (Dow et al., 2005b).
- International VLBI Service (IVS) (Schlüter et al., 2002).
- International Laser Ranging Service (ILRS) (Pearlman et al., 2002).
- International DORIS Service (IDS) (Tavernier et al., 2005).

These scientific services, as well as gravitational field services and an expected future altimetry service, are integral components of the future 'Global Geodetic Observing System' (GGOS) (Rummel et al., 2005; http://www.ggos.org). Closer cooperation and understanding through GGOS are expected to bring significant improvements to the ITRF and to scientific uses of geodesy in general (Dow et al., 2005a).

Figure 4 shows the current status of colocated space geodetic sites, which forms the foundation for ITRF and GGOS. Colocation is essential to exploit the synergy of the various techniques, and so increasing the number and quality of colocated sites will be a high priority for GGOS. In addition, the IGS adapts to incorporate new GNSS systems as they come online.

3.11.2 GPS and Basic Principles

3.11.2.1 Basic Principles

GPS positioning is based on the principle of 'trilateration,' which is the method of determining position by measuring distances to points of known positions (not to be confused with triangulation, which measures angles between known points). At a minimum, trilateration requires three ranges to three known points. In the case of GPS, the known points would be the positions of the satellites in view. The measured ranges would be the distances between the GPS satellites and a user's GPS receiver. (Note that GPS is a completely passive system from which users only receive signals.) GPS receivers, on the other hand, cannot measure ranges directly, but rather 'pseudoranges.' A pseudorange is a measurement of the difference in time between the receiver's local clock and an atomic clock on board a satellite. The measurement is multiplied by the speed of light to convert it into units of range (meters):

$$\text{pseudorange} = (\text{receiver\_time} - \text{satellite\_time}) \times \text{speed\_of\_light} \quad [1]$$

The satellite effectively sends its clock time by an encoded microwave signal to a user's receiver. It does this by multiplying a sinusoidal carrier wave by a known sequence ('code') of +1 and −1, where the timing of the signal (both code and carrier wave) is controlled by the satellite clock. The receiver generates an identical replica code and then performs a cross

Table 2 IGS (and broadcast) product availability and quality

		Accuracy[a]	*Latency*	*Updates*	*Interval*
GPS satellite ephemerides					
Broadcast[b]	Orbits	160 cm	Real time	N/A	N/A
	Satellite clocks	7 ns			
Ultrarapid (predicted half)	Orbits	10 cm	Real time	6 h	15 min
	Satellite clocks	5 ns			
Ultrarapid (observed half)	Orbits	<5 cm	3 h	6 h	15 min
	Satellite clocks	0.2 ns			
Rapid	Orbits	<5 cm	17 h	Daily	15 min
	All clocks	0.1 ns			5 min
Final	Orbits[c]	<5 cm	13 days	Weekly	15 min
	All clocks[d]	0.1 ns			5 min
GLONASS satellite ephemerides					
Final		15 cm	2 weeks	Weekly	15 min
IGS station coordinates[e]					
Positions	Horizontal	3 mm	12 days	Weekly	Weekly
	Vertical	6 mm			
Velocities	Horizontal	2 mm $year^{-1}$	12 days	Weekly	~Years
	Vertical	3 mm $year^{-1}$			
Earth rotation parameters[f]					
Ultrarapid (predicted half)	Pole position	0.3 mas	Real time	6 h	6 h
	Pole rate	0.5 mas day^{-1}			
	Length of day	0.06 ms			
Ultrarapid (observed half)	Pole position	0.1 mas	3 h	6 h	6 h
	Pole rate	0.3 mas day^{-1}			
	Length of day	0.03 ms			
Rapid	Pole position	<0.1 mas	17 h	Daily	Daily
	Pole rate	<0.2 mas day^{-1}			
	Length of day	0.03 ms			
Final	Pole position	0.05 mas	13 days	Weekly	Daily
	Pole rate	<0.2 mas day^{-1}			
	Length of day	0.02 ms			

[a]Generally, precision (based on scatter of solutions) is better than the accuracy (based on comparison with independent methods).
[b]Broadcast ephemerides only shown for comparison (but are also available from IGS).
[c]Orbit accuracy based on comparison with satellite laser ranging to satellites.
[d]Clock accuracy is expressed relative to the IGS timescale, which is linearly aligned to GPS time in 1-day segments.
[e]Station coordinate and velocity accuracy based on intercomparison statistics from ITRF.
[f]Earth rotation parameters based on intercomparison statistics by IERS. IGS uses VLBI results from IERS Bulletin A to calibrate for long-term LOD biases.
Source: IGS Central Bureau, http://igscb.jpl.nasa.gov.courtesy of A. Moore.

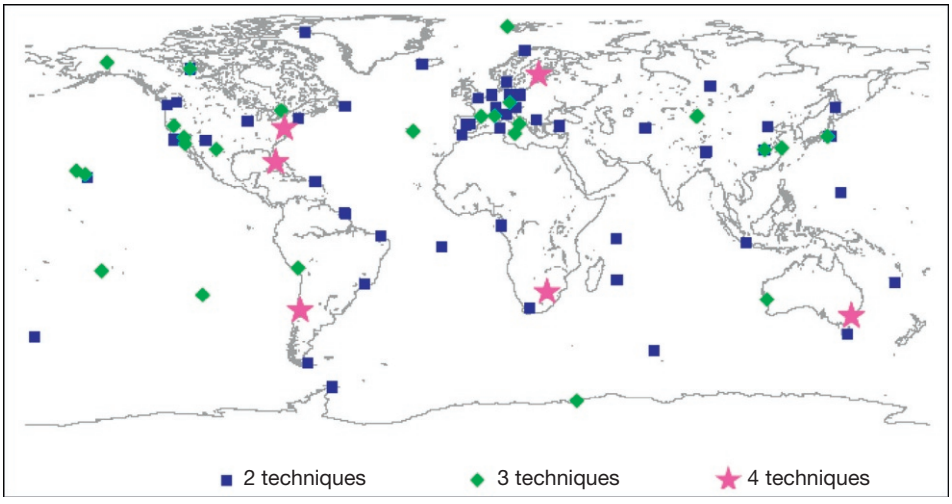


Figure 4 Distribution of colocated space geodetic stations that have at least two different operational techniques of GPS, VLBI, SLR, and DORIS. Courtesy of Z. Altamimi.

correlation with the incoming signal to compute the required time shift to align the codes. This time shift multiplied by the speed of light gives the pseudorange measurement.

The reason the measurement is called a 'pseudorange' is that the range is biased by error in the receiver's clock (typically a quartz oscillator). However, this bias at any given time is the same for all observed satellites, and so it can be estimated as one extra parameter in the positioning solution. There are also (much smaller) errors in the satellites' atomic clocks, but GPS satellites handle this by transmitting another code that tells the receiver the error in its clock (which is routinely monitored and updated by the US Department of Defense).

Putting all this together, point positioning with GPS therefore requires pseudorange measurements to at least four satellites, where information on the satellite positions and clocks is also provided as part of the GPS signal. Three coordinates of the receiver's position can then be estimated simultaneously along with the receiver's clock offset. By this method, GPS positioning with few meter accuracy can be achieved by a relatively low-cost receiver.

Hence, GPS also allows the user to synchronize time to the globally accessible atomic standard provided by GPS. In fact, the GPS atomic clocks form part of the global clock ensemble that define Universal Coordinated Time (UTC). Note that since GPS time began (6 January 1980), there have accumulated a number of leap seconds (16 s as of July 2012) between GPS time (a continuous timescale) and UTC (which jumps occasionally to maintain approximate alignment with the variable rotation of the Earth). Synchronization to GPS time (or UTC) can be achieved to <0.1 μs using a relatively low-cost receiver. This method is suitable for many time-tagging applications, such as in seismology, in SLR, and even for GPS receivers themselves. That is, by using onboard point positioning software, GPS receivers can steer their own quartz oscillator clocks through a feedback mechanism such that observations are made within a certain tolerance of GPS time.

A fundamental principle to keep in mind is that GPS is a timing system. By the use of precise timing information on radio waves transmitted from the GPS satellite, the user's receiver can measure the range to each satellite in view and hence calculate its position. Positions can be calculated at every measurement epoch, which may be once per second when applied to car navigation (and in principle as frequently as 50 Hz). Kinematic parameters such as velocity and acceleration are secondary, in that they are calculated from the measured time series of positions.

3.11.2.2 GPS Design and Consequences

The GPS has three distinct segments:

1. The space segment, which includes the constellation of ~30 GPS satellites that transmit the signals from space down to the user, including signals that enable a user's receiver to measure the biased range (pseudorange) to each satellite in view and signals that tell the receiver the current satellite positions, the current error in the satellite clock, and other information that can be used to compute the receiver's position.
2. The control segment (in the US Department of Defense), which is responsible for the monitoring and operation of the space segment, including the uploading of information that can predict the GPS satellite orbits and clock errors into the near future, which the space segment can then transmit down to the user.
3. The user segment, which includes the user's GPS hardware (receivers and antennas) and GPS data processing software for various applications, including surveying, navigation, and timing applications.

The satellite constellation is designed to have at least four satellites in view anywhere, anytime, to a user on the ground. For this purpose, there are nominally 24 GPS satellites distributed in six orbital planes. In addition, there is typically an active spare satellite in each orbital plane, bringing the total number of satellites closer to 30. The orientation of the satellites is always changing, such that the solar panels face the sun and the antennas face the center of the Earth. Signals are transmitted and received by the satellite using microwaves. Signals are transmitted to the user segment at frequencies L1 = 1575.42 MHz and L2 = 1227.60 MHz in the direction of the Earth. This signal is encoded with the 'navigation message,' which can be read by the user's GPS receiver. The navigation message includes orbit parameters (often called the 'broadcast ephemeris'), from which the receiver can compute satellite coordinates (X,Y,Z). These are Cartesian coordinates in a geocentric system, known as WGS 84, which has its origin at the Earth's center of mass, Z-axis pointing toward the North Pole, X pointing toward the prime meridian (which crosses Greenwich), and Y at right angles to X and Z to form a right-handed orthogonal coordinate system. The algorithm that transforms the orbit parameters into WGS 84 satellite coordinates at any specified time is called the 'Ephemeris Algorithm' (e.g., Leick, 2004). For geodetic purposes, precise orbit information is available over the Internet from civilian organization such as the IGS in the Earth-fixed reference frame.

According to Kepler's laws of orbital motion, each orbit takes the approximate shape of an ellipse, with the Earth's center of mass at the focus of the ellipse (Figure 5). For a GPS orbit, the

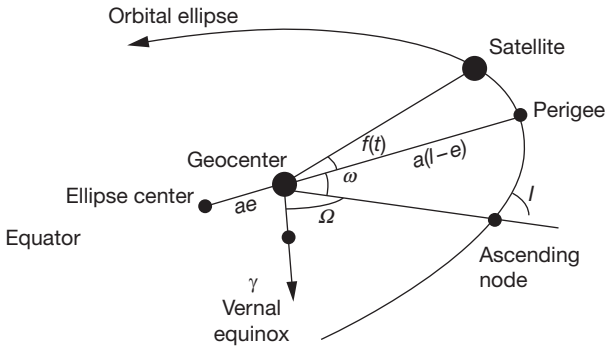


Figure 5 Diagram illustrating the Keplerian orbital elements: semimajor axis *a*, eccentricity *e*, inclination I, argument of perigee (closest approach) ω, right ascension of the ascending node Ω, and true anomaly *f* as a function of time *t*. The geocenter is the Earth's center of mass; hence, satellite geodesy can realize the physical origin of the terrestrial reference system. This diagram is exaggerated, as GPS orbits are almost circular.

eccentricity of the ellipse is so small (0.02) that it is almost circular. The semimajor axis (largest radius) of the ellipse is approximately 26 600 km, or approximately 4 Earth radii.

The six orbital planes rise over the equator at an inclination angle of 55°. The point at which they rise from the Southern to Northern Hemisphere across the equator is called the 'right ascension of the ascending node.' Since the orbital planes are evenly distributed, the angle between the six ascending nodes is 60°.

Each orbital plane nominally contains four satellites, which are generally not spaced evenly around the ellipse. Therefore, the angle of the satellite within its own orbital plane, the 'true anomaly,' is only approximately spaced by 90°. The true anomaly is measured from the point of closest approach to the Earth (the perigee). Instead of specifying the satellite's anomaly at every relevant time, it is equivalent to specify the time that the satellite had passed perigee and then compute the satellite's future position based on the known laws of motion of the satellite around an ellipse. Finally, the argument of perigee specifies the angle between the equator and perigee. Since the orbit is nearly circular, this orbital parameter is not well defined, and alternative parameterization schemes are often used.

Taken together (the eccentricity, semimajor axis, inclination, right ascension of the ascending node, time of perigee passing, and argument of perigee), these six parameters define the satellite orbit (according to the Keplerian model). These parameters are known as Keplerian elements. Given the Keplerian elements and the current time, it is possible to calculate the coordinates of the satellite.

However, GPS satellites do not move in perfect ellipses, so additional parameters are necessary. Nevertheless, GPS does use Kepler's laws to its advantage, and the orbits are described in the broadcast ephemeris by parameters that are Keplerian in appearance. Additional parameters must be added to account for non-Keplerian behavior. Even this set of parameters has to be updated by the control segment every hour for them to remain sufficiently valid.

Several consequences of the orbit design can be deduced from the previously mentioned orbital parameters and Kepler's laws of motion. First of all, the satellite speed is $\sim$4 km s^{-1} relative to the Earth's center. All the GPS satellite orbits are prograde, which means the satellites move in the direction of the Earth's rotation. Therefore, the relative motion between the satellite and a user on the ground must be less than 4 km s^{-1}. Typical values around 1 km s^{-1} can be expected for the relative speed along the line of sight (range rate).

The second consequence is the phenomena of 'repeating ground tracks' every day. The orbital period is approximately $T = 11$ h 58 min; therefore, a GPS satellite completes two revolutions in 23 h 56 min. This is intentional, as it equals one sidereal day, the time it takes for the Earth to rotate 360°. Therefore, every day (minus 4 min), the satellite appears over the same geographic location on the Earth's surface. The 'ground track' is the locus of points on the Earth's surface that is traced out by a line connecting the satellite to the center of the Earth. The ground track is said to repeat. From the user's point of view, the same satellite appears in the same direction in the sky every day minus 4 min. Likewise, the 'sky tracks' repeat.

So from the point of view of a ground user, the entire satellite geometry repeats every sidereal day. Consequently, any errors correlated with satellite geometry will repeat from one day to the next. An example of an error tied to satellite geometry is 'multipath,' which is due to the antenna also sensing signals from the satellite that reflect and refract from nearby objects. In fact, it can be verified that, because of multipath, observation residuals do have a pattern that repeats every sidereal day. Therefore, such errors will not significantly affect the repeatability of coordinates estimated each day. However, the accuracy can be significantly worse than the apparent precision for this reason.

Another consequence of this is that the same subset of the 24 satellites will be observed every day by someone at a fixed geographic location. Generally, not all 24 satellites will be seen by a user at a fixed location. This is one reason why there needs to be a global distribution of receivers around the globe to be sure that every satellite is tracked sufficiently well.

The inclination angle of 55° also has consequences for the user. Note that a satellite with an inclination angle of 90° would orbit directly over the poles. Any other inclination angle would result in the satellite never passing over the poles. From the user's point of view, the satellite's sky track would never cross over the position of the celestial pole in the sky. In fact, there would be a 'hole' in the sky around the celestial pole where the satellite could never pass. For a satellite constellation with an inclination angle of 55°, there would therefore be a circle of radius at least 35° around the celestial pole, through which the sky tracks would never cross. This has a big effect on the satellite geometry as viewed from different latitudes. An observer at the pole would never see a GPS satellite rise above 55° elevation. Most of the satellites would hover close to the horizon. Therefore, vertical positioning is slightly degraded near the poles. An observer at the equator would see some of the satellites passing overhead but would tend to deviate away from points on the horizon directly to the north and south.

Due to a combination of the Earth's rotation and the fact that the GPS satellites are moving faster than the rotation of the Earth, the satellites actually appear to move approximately north–south or south–north to an observer at the equator, with very little east–west motion. Therefore, the closer the observer is to the equator, the better determined the north component of relative position becomes as compared with the east component. An observer at midlatitudes in the Northern Hemisphere would see satellites anywhere in the sky to the south, but there would be a large void toward the north. This has consequences for site selection, where a good view is desirable to the south, and the view to the north is less critical. For example, one might want to select a site in the Northern Hemisphere that is on a south-facing slope (and vice versa for an observer in the Southern Hemisphere).

3.11.2.3 Introducing High-Precision GPS

By measuring pseudoranges to at least four satellites with relatively low-cost equipment, GPS can readily provide users with a positioning accuracy of meters and a timing accuracy of 0.1 μs. On the other hand, geodetic GPS positioning with an accuracy of a few millimeters requires a number of significant improvements to the technique described in the preceding text, which will be emphasized in this section. For example,

accurate positioning requires accurate knowledge of the GPS satellite positions and satellite clock offsets. For standard GPS positioning, this 'ephemeris' information is broadcast by the GPS satellites in the so-called navigation message; however, it is not sufficiently accurate for geodetic applications.

In addition to the three GPS segments listed earlier, one could informally include the 'service segment' consisting of civilian networks that provide the user segment with data and services to enhance positioning accuracy. This information can be transmitted to the user in a variety of ways, such as by the Internet, cell phone, and geostationary satellite. The part of this service segment that is relevant to geodetic positioning would be the International GNSS Service (IGS), an international collaboration of geodesists that provides high-accuracy data on satellite orbits and clocks. IGS also provides data from reference stations around the globe, at accurately known coordinates that account for plate tectonics and other geophysical movements such as earthquakes. Thus, the IGS enables users to position their receivers anywhere on the globe with an accuracy of millimeters in a consistent terrestrial reference frame. But this is only solves one of the many problems toward achieving geodetic precision.

In practice, high-precision GPS geodesy requires a minimum of five satellites in view, because it is essential to estimate parameters to model tropospheric refraction. At an absolute minimum, one zenith delay is estimated, which can be mapped to delay at any elevation angle using a 'mapping function' based on tropospheric models.

Geodetic applications require much more sophisticated GPS receivers that measure not only the pseudorange observable but also the so-called 'carrier phase' observable. The carrier phase observable is the difference between (1) the phase of the incoming carrier wave (upon which the codes are transmitted) and (2) the phase of a signal internally generated by the receiver, which is synchronized with the receiver clock. When multiplied by the ~20 cm wavelength of the carrier wave, the result is a biased distance to the satellite. Indeed, this is a type of pseudorange that is about 100 times more precise than the coded pseudoranges. The downside to the carrier phase observable is that in addition to the receiver clock bias, there is an additional bias of an unknown number of wavelengths. It is possible to resolve this bias exactly by the so-called 'ambiguity resolution' techniques. Ambiguity resolution is essential to achieve the highest possible precision for geodetic applications. Hence, in units of range, the observed carrier phase can be expressed as

$$\text{carrier\_phase} = (\text{reference\_phase} - \text{signal\_phase} + \text{integer}) \times \text{carrier\_wavelength} \quad [2]$$

Note that the signal phase is generated by the satellite clock and that the reference phase is generated by the receiver clock; hence, eqn [2] is just a very precise form of eqn [1] for the pseudorange, except that it has an integer–wavelength ambiguity. (In fact, this is why the sign of the phase difference was chosen by subtracting the incoming signal phase from the reference phase.) Therefore, the observable models for eqns [1] and [2] are very similar and relate to the theoretical difference between the reading of the receiver clock (time of reception) and the satellite clock (time of transmission), including clock biases.

This similarity of models has enabled the development of automatic signal processing algorithms to check the integrity of the data, such as the detection of data outliers and jumps in the integer ambiguity (the so-called cycle slips), which occur when the receiver loses lock on the signal, for example, due to a temporary obstruction between the ground antenna and the satellite. In fact, the pseudorange data can be used together with the carrier phase data to correct for the initial integer ambiguity (Blewitt, 1989) and for subsequent cycle slips (Blewitt, 1990).

For geodetic positioning, both pseudoranges and carrier phases are measured at two different frequencies (L1 at 19.0 cm wavelength and L2 at 24.4 cm), to provide a self-calibration of delay in the Earth's ionosphere. So in total, there are four observations that are fundamental to high-precision GPS geodesy: two pseudoranges and two carrier phases. This enables more algorithms to assure the integrity of the data and allows for monitoring of the ionosphere itself.

Another requirement for geodetic positioning is the use of highly specialized stochastic multiparameter estimation software by modeling the carrier phase data, including modeling of the satellite–station geometry, Earth's atmosphere, solid Earth tides, Earth's rotation, antenna effects, circular polarization effects (phase windup), and relativistic effects (both special and general). In addition, the software must be capable of detecting and correcting integer offsets in the carrier phase observables (cycle slips) and must be capable of resolving the integer ambiguity in the initial phase measurements.

In summary, therefore, geodetic GPS requires

- geodetic-class GPS receivers capable of acquiring dual-frequency carrier phase data;
- geodetic-class satellite orbit and clock information, which is available from the IGS;
- simultaneous observations to a minimum of five satellites;
- specialized postprocessing software (not on the receiver itself) that embodies high-accuracy observable models, carrier phase data processing algorithms, and simultaneous parameter estimation.

The quality of the IGS orbit and clock data depends on their latency, so generally, there is a trade-off between latency and accuracy. Currently, the ultrarapid IGS product is actually a prediction from 3 to 9 h ago. Even though there are atomic clocks on board the GPS satellites, the clock time is much more difficult to predict than the satellite orbits. In the case that sufficiently accurate clock data are not yet available, it is nevertheless possible to produce geodetic-class solutions for relative positions between ground stations. This is achieved either by (1) solving for satellite clock biases at every epoch as part of the positioning solution or equivalently by (2) differencing data between ground stations to cancel out the clock bias. Furthermore, data can be differenced again ('double difference') between satellites to cancel out the receiver clock bias rather than estimate it as a parameter (**Figure 6**).

In practice, the following different approaches to estimating positions all give results that are of geodetic quality (with errors measured in millimeters) and typically agree very well:

- Precise point positioning (PPP) of single stations using precise orbit and clock data.

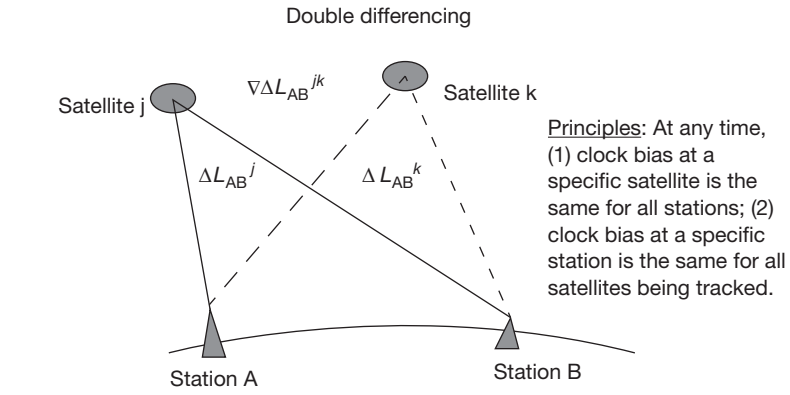

Double differencing
Satellite j
$\nabla\Delta L_{AB}{}^{jk}$
Satellite k
$\Delta L_{AB}{}^{j}$
$\Delta L_{AB}{}^{k}$
Principles: At any time, (1) clock bias at a specific satellite is the same for all stations; (2) clock bias at a specific station is the same for all satellites being tracked.
Station A
Station B
At each epoch, difference the single difference data between satellites:
$\nabla\Delta L_{AB}{}^{jk} = \Delta L_{AB}{}^{j} - \Delta L_{AB}{}^{k}$

Figure 6 Diagram illustrating double differencing of GPS data. The idea is to differentiate away the satellite and station clock biases. Double differencing is equivalent to estimating the clock biases explicitly when processing undifferenced data.

- Relative positioning of networks by clock estimation, using precise orbit data.
- Relative positioning of networks by double-difference data (Figure 6), using precise orbit data.

All of these three methods are in common use today for geophysical research purposes. In each case, dual-frequency pseudorange and carrier phase data types are used.

3.11.2.4 GPS Observable Modeling

This section describes how GPS observables are typically modeled by geodetic-quality software packages. First, however, a few more specific details on the GPS signals are required. The signals from a GPS satellite are fundamentally driven by an atomic clock precisely at frequency 10.23 MHz. Two sinusoidal carrier signals are generated from this signal by multiplying the frequency by 154 for the L1 channel (frequency = 1575.42 MHz; wavelength = 19.0 cm) and 120 for the L2 channel (frequency = 1227.60 MHz; wavelength = 24.4 cm). Information is encoded in the form of binary bits on the carrier signals by a process known as phase modulation. The binary digits 0 and 1 are actually represented by multiplying the electrical signals by either +1 or −1.

For purposes of observable modeling, here, the observables (all in units of meters) will be called L1 and L2 for the two types of carrier phase and P1 and P2 for the two types of pseudorange. The actual observable types are numerous due to different methods of correlating the signals; however, the fundamental observation equations can be written in the same generic way, with the exception that there is generally a bias associated with each observable types, including instrumental bias and, in the case of the carrier phase, an integer wavelength bias. (For some older signal-squaring receivers, the bias is a half-integer wavelength.) Here, it is simply assumed that such biases are not problematic, which is typically the case, and so the interobservable biases are not explicitly modeled.

Taking eqns [1] and [2], the generic (pseudorange or carrier phase) GPS observation P_j^i at receiver j (subscript, on the ground) from satellite i (superscript, up in space) can be modeled:

$$P_j^i = c\left(T_j - \overline{T}^i\right) + B_j^i \qquad [3]$$

Since special relativity and general relativity prove to be important in the model, care must be taken to define each term with respect to a reference frame. Thus, T_j is the time according to the receiver clock coincident with signal reception (used as the time tag, recorded with the observation), $\overline{T}^i$ is the time according to the satellite clock coincident with signal transmission (which imprints its signature on the signal, hence the bar, which denotes time local to the satellite), B_j^i is a frame-invariant bias associated with this type of observation, and c is the frame-invariant speed of light in a vacuum. In addition, this observation is recorded at an epoch with time tag T_j (the same for all satellites observed at that epoch).

The clock difference can be rewritten as the sum of four time differences:

$$P_j^i = c\left\{\left(T_j - t_j\right) + \left(t_j - t^i\right) + \left(t^i - \overline{t}^i\right) + \left(\overline{t}^i - \overline{T}^i\right)\right\} + B_j^i \qquad [4]$$

where t_j is the coordinate time at the receiver, t^i is the coordinate time at the satellite, and $\overline{t}^i$ is the proper time at the satellite (the time kept by a perfect clock on board the satellite). 'Coordinate time' simply means the timescale that is actually used to compute the models. It is convenient to take coordinate time in the 'local Earth' frame (that of a perfect clock on the geoid) (Ashby and Allan, 1984). Appropriate timescales for this purpose include Terrestrial Dynamic Time (TDT) and International Atomic Time (TIA), but for the discussion here, it is convenient to choose GPS time. The important thing to keep in mind (to cut through the confusion of all these conventions) is that all these timescales ideally run at the same rate as UTC, with the unit of time being the SI second (Kaplan, 1981), and so all these scales only differ by conventional constant offsets (and leap seconds).

The four time-difference terms found in eqn [4] can be written as follows. First, the difference in receiver clock time and coordinate time is simply the receiver clock bias, which we will model as an independent parameter τ_j at every epoch (at every value of T_j):

$$(T_j - t_j) = \tau_j \qquad [5]$$

The clock bias includes the sum of a clock error (with respect to proper time) plus a minor relativistic bias due to the geodetic location of the receiver clock.

The second term is the difference between coordinate time at the receiver and satellite, the so-called light-time equation (here expressed as a range):

$$\begin{aligned} c(t_j - t^i) &= r_j^i + \sum_{\text{prop}} \Delta r_{\text{prop}_j}^{\ i} \\ &= \left|\mathbf{r}_j(t_j) - \mathbf{r}^i(t^i)\right| + \Delta r_{\text{GR}_j}^{\ i} + \Delta r_{\text{ion}_j}^{\ i} + \Delta r_{\text{trop}_j}^{\ i} \\ &\quad + \Delta r_{\text{pcv}_j}^{\ i} + \Delta r_{\text{circ}_j}^{\ i} + \cdots \end{aligned} \qquad [6]$$

where r_j^i is the Euclidean distance between the satellite and receiver and $\Delta r_{\text{prop}j}^i$ represent various propagation delays, which are a function of satellite–station geometry, arising from space–time curvature (general relativity); ionosphere, troposphere, and antenna phase center variations; circular polarization effects; and other propagation terms as necessary. In eqn [6], $\mathbf{r}_j$ is the geocentric receiver position at the time of reception and $\mathbf{r}^i$ is the geocentric satellite position at the time of transmission. The reference frame for the light-time equation is taken to be J2000, the conventional Earth-centered inertial (ECI) frame (so the axes do not corotate with the Earth), because the speed of light is a constant in all directions in an inertial frame. In addition, the ECI frame is most convenient for integrating the satellite equations of motion, if the goal is to estimate satellite orbits. However, it can be more convenient to solve the light-time equation in a rotating, Earth-centered, Earth-fixed (ECEF) reference frame, given that satellite orbit positions and velocities are typically distributed in the ECEF frame. In this case, the equation would also need to account for the Sagnac effect (e.g., Ashby, 2004), which accounts for the effect of the Earth's rotation on light propagation when viewed in the ECEF frame. A simple way to handle this is to consider the ECEF and ECI frames as being instantaneously aligned at reception time. Thus, all that is required is to compute the satellite velocity in the ECI frame by adding the velocity due to the Earth's rotation at the satellite to the ECEF velocity and then compute the ECI position of the satellite at transmit time using this ECI velocity. Thus, the Sagnac effect can be considered simply as a correction for the Earth's rotation.

The general relativistic delay can be computed as

$$\Delta r_{\text{GR}_j}^{\ i} = \frac{2GM_\oplus}{c^2} \ln \frac{r_j + r^i + r_i^j}{r_j + r^i - r_i^j} \qquad [7]$$

where $GM_\oplus$ is the Earth's gravitational constant and, in general, $r \equiv |\mathbf{r}|$. This is sometimes called the Shapiro delay after Shapiro (1964) and is typically at the 10 mm level for signals from satellites in medium Earth orbit. It accounts for the distance traveled, being longer than the Euclidean distance, caused by space–time curvature in the Earth's gravitational field. Equation [7] is a sufficiently good approximation, derived by integrating the effect of the Earth's gravitational potential along a straight light path; so the term 'space–time curvature' should not be confused with gravitational bending of the direction of light, which is negligible. Rather, the curvature can be thought of primarily as the 'stretching' of space along the direction of travel.

Antenna effects such as phase center variation $\Delta r_{\text{pcv}j}^i$ [Schupler et al., 1994] and circular polarization $\Delta r_{\text{circ}j}^i$ ('phase windup') (Wu et al., 1993) are examples of important effects that have been researched and applied to improve positioning accuracy, but as nongeophysical effects, they are beyond the scope of this text. Antenna calibrations for both the GPS satellite and station antennas are available from the IGS and should be applied (Schmid et al., 2005).

Ionospheric delay can be adequately modeled as being inversely proportional to the squared frequency f of the carrier wave:

$$\Delta r_{\text{ion}_j}^{\ i}(f) = \pm k \frac{\text{TEC}_j^i}{f^2} \qquad [8]$$

where the positive sign is taken for pseudoranges and the negative sign for carrier phase observations. The term TEC refers to 'total electron content,' which is excited by solar radiation and so is highly variable through the day and is sensitive to geographic location. The constant k can be derived from the theory of electromagnetic (EM) wave propagation in plasmas. Delays at GPS frequencies can be as large as 100 m near the equator, peaking around 2 pm local time, and can be as small as centimeters at midlatitudes between midnight and dawn. In contrast, higher-order terms are of the order of millimeters and are typically ignored, although including them in the model is one of many themes of current research (Kedar et al., 2003). An appropriate linear combination of observations eliminates the frequency-squared term exactly, leaving all nondispersive terms in the model unchanged. In fact, the 'ionosphere-free' combination of carrier phases can be so defined (and similarly for the pseudoranges):

$$\begin{aligned} \text{LC} &= \frac{{f_1}^2 L1 - {f_2}^2 L2}{({f_1}^2 - {f_2}^2)} \\ &\cong 2.546 L1 - 1.546 L2 \end{aligned} \qquad [9]$$

Hence, k and TEC are not explicitly needed to compute the ionosphere-free data. The coefficients in the preceding text can be computed exactly by substituting $f_1 = 154$ and $f_2 = 120$, owing to the properties of the GPS signals described at the beginning of this section. As an aside, if the ionosphere is the geophysical scientific target of interest, then differencing the observations at two different frequencies results in a 'geometry-free' observation from which TEC can be estimated:

$$\begin{aligned} PI &= P1 - P2 \\ &= k\text{TEC}\left(\frac{1}{{f_1}^2} - \frac{1}{{f_2}^2}\right) + \text{bias} \end{aligned} \qquad [10]$$

Using GPS stations located around the globe, this method is now routinely used to map ionospheric TEC. A side benefit of this method is the estimation of the interchannel bias between observables at L1 and L2 frequency, which can be monitored for long-term variability and used as input to ambiguity resolution algorithms.

The tropospheric delay (a fraction of which actually occurs in the stratosphere) is almost entirely nondispersive

(independent of frequency) at GPS L-band frequencies and so must be handled in a different way. Whereas it is possible in principle to model tropospheric delay based on ground-based meteorologic observations, in practice, this has not proven to be sufficiently accurate. The key to successful tropospheric modeling is the estimation of the delay at zenith, by accurately modeling the relationship between zenith delay Z and delay at lower elevations ε, for example,

$$\Delta r_{\mathrm{trop}_j}^{\ i} = \frac{Z_j}{\sin \varepsilon_j^i} \qquad [11]$$

where the inverse sign of elevation angle is the simplest example of a 'mapping function,' which can be derived by assuming a horizontally layered troposphere over a flat Earth. This model breaks down rapidly for $\varepsilon < 20°$. More accurate modeling (Truehaft and Lanyi, 1987) requires modifying the mapping function to account for Earth curvature and partitioning the delay into a so-called dry and wet components, which have different characteristic scale heights (~10 km and ~2 km, respectively):

$$\begin{aligned} \Delta r_{\mathrm{trop}_j}^{\ i} &= \Delta r_{\mathrm{dry}_j}^{\ i} + \Delta r_{\mathrm{wet}_j}^{\ i} \\ &= Z_{\mathrm{dry}} F_{\mathrm{dry}}\left(\varepsilon_j^i\right) + Z_{\mathrm{wet}} F_{\mathrm{wet}}\left(\varepsilon_j^i\right) \end{aligned} \qquad [12]$$

Due to the inherent weakness in the determination of height with both GPS and VLBI, accurate modeling of mapping functions has always been and remains an active area of research. The wet delay is caused by the interaction of the EM wave with the static dipole of molecular water. The dry delay is due to the dynamic dipole induced by the EM wave on all component molecules in the atmosphere, including a (small) contribution from water (and so 'dry' is just a conventional, perhaps, misleading term). Typical values for the dry and wet delay are 2.1 and 0.1 m, respectively, to within ~10 cm.

The dry component can be adequately modeled as a function of hydrostatic pressure at the altitude of the receiver. Nominal values can be computed in the absence of meteorologic data by assuming a nominal surface pressure at sea level and then subtracting a correction for altitude, assuming that pressure decays exponentially with altitude. The wet component is typically assumed to have a nominal value of zero, and Z_{wet} is then estimated from the GPS data along with the positioning solution. Note that in this case, the estimated value of Z_{wet} would absorb most (but not all) of the obvious inadequacies of the nominal model for Z_{dry}. Whereas this is currently the standard method in high-precision GPS geodesy, the limitations of this approach are an active area of research (Tregoning and Herring, 2006).

The tropospheric delay model is important not only for solving for geodetic position but also for the study of the troposphere itself. For this application, estimates of troposphere delay can be converted into precipitable water vapor in the atmosphere, which can then be used as input for weather forecasting and climate modeling. For this application, surface meteorologic data are essential to more accurately partition the dry and wet components of delay (Bevis et al., 1992).

Now, returning to the light-time equation, even if we had perfect propagation models, the light-time equation needs to be solved iteratively rather than simply computed, because at first, we do not have a nominal value for t^i, the coordinate time of signal transmission. The procedure is as follows:

- Starting with the observation time tag T_j, use eqn [5] and a nominal value for the receiver clock bias τ_j (which may be zero or a preliminary estimate) to compute the coordinate time of signal reception t_j. Note that the assumed clock bias affects the subsequent computation of geometric range, indicating the need for iterative estimation. (This problem can be more conveniently addressed by accounting for the range rate in the partial derivative with respect to the receiver clock parameter.)
- Given a modeled station position $\mathbf{r}_j$ at time t_j, and an interpolated table of modeled satellite positions $\mathbf{r}^i$ as a function of coordinate time t^i, iteratively compute the coordinate time of transmission using

$$ct^j[n+1] = ct^j[n] + \frac{ct_j - ct^j[n] - r_j^i[n] - \sum_{\mathrm{prop}} \Delta r_{\mathrm{prop}_j}^{\ i}[n]}{1 - \dot{\mathbf{r}}^i[n] \cdot \hat{\mathbf{r}}_j^i[n]/c} \qquad [13]$$

where $\hat{\mathbf{r}}_j^i = \mathbf{r}_j^i / r_j^i$ are the direction cosines to the receiver from the satellite. It is to be understood that at the nth iteration, for example, the satellite position $\mathbf{r}^i$ and ECI velocity $\dot{\mathbf{r}}^i$ are both interpolated to time $t^i[n]$. By virtue of this equation converging very quickly, it is sufficient to initialize the transmission time to the reception time $t^i[0] = t_j$.

Being in the ECI frame (J2000), the receiver position must account for the Earth's rotation and geophysical movements of the Earth's surface:

$$\mathbf{r}_j(t_j) = \mathbf{PNUXY}\left[\mathbf{x}_{0j} + \sum\nolimits_k \Delta \mathbf{x}_{kj}(t_j)\right] \qquad [14]$$

Here, **PNUXY** is the multiple of 3×3 rotation matrices that account (respectively) for precession, nutation, rate of rotation, and polar motion (in two directions). The bracketed term represents the receiver position in the (corotating) conventional Earth-fixed terrestrial reference frame known as ITRF. Conventional station position $\mathbf{x}_{0j}$ is specified by station coordinates in ITRF at some conventional epoch, and $\Delta \mathbf{x}_{kj}(t_j)$ represents the displacement from the epoch position due to geophysical process k, for example, accounting for the effects of plate tectonics, solid Earth tides, etc. Equation [14] together with [6] forms the fundamental basis of using GPS as a geophysical tool, and this will be explored later.

Returning now to the original observation eqn [4], the third term is the difference between coordinate time and proper time at the satellite. According to special relativity, GPS satellite clocks run slow relative to an observer on the Earth's surface due to relative motion. In contrast, general relativity predicts that the satellite clocks will appear to run faster when observed from the Earth's surface due to the photons gaining energy (gravitational blue shift) as they fall into a gravitational well. These two effects do not entirely cancel. With appropriate foresight in the design of GPS, the satellite clocks operate at a slightly lower frequency than 10.23 MHz to account for both special relativity and general relativity, so that their frequency would be 10.23 MHz (on average) as viewed on the Earth's surface. The residual relativistic term can be computed from

$$\left(t^i - \bar{t}^i\right) = 2\mathbf{r}^i\left(t^i\right) \cdot \dot{\mathbf{r}}^i\left(t^i\right)/c^2 \qquad [15]$$

It is inconsequential as to whether eqn [15] is computed in the ECI or ECEF frame. This result assumes an elliptical orbit

about a point (or spherically layered) mass and assumes that additional relativistic effects are negligible. The expected accuracy is at the level of 10^{-12}, comparable with the level of stability of the satellite atomic clocks.

The fourth term is the difference between proper time and clock time at the satellite, which is simply the negative error in the satellite clock:

$$\left(\overline{t}^i - \overline{T}^i\right) = -\tau^i \qquad [16]$$

Unlike errors in the station clock, errors in the satellite clock do not affect the model of geometric range (in the light-time equation) and so are not required in advance to compute that part of the model. However, they do affect the observable itself, and so not accounting for satellite clock error would result in an error in estimated receiver position. Despite the satellite clocks being atomic, they are not sufficiently stable to be predicted forward in time for geodetic applications (though this is the method for standard positioning with GPS). Therefore, this term is estimated independently at every epoch as part of the positioning solution (or alternatively, observations equations can be differenced between pairs of observing receivers to eliminate this parameter). As a consequence, geodetic-quality PPP of individual stations in real time presents a challenge that is the topic of current research.

Finally, each observation type has an associated bias. Typically, receivers are designed so that these biases either should be calibrated or are stable in time. Like the case for the satellite clock error, these biases have no effect on the computed geometric range and so do not need to be known in advance; however, they are present in the observations themselves and so can affect positioning accuracy unless they are absorbed by parameters in the least-squares solution. It turns out for the most part that biases between observable types can be ignored for purposes of positioning, because they can be absorbed into the station or satellite clock bias parameters as part of the least-squares positioning solution. For purposes of accurate timing, however, special considerations are required to calibrate such biases. Some of the inter-observable biases are monitored by major GPS analysis centers and made routinely available if needed.

The most important bias to consider for geodetic applications is the carrier phase bias, for which its double difference has an integer ambiguity (Blewitt, 1989). The carrier phase bias is not predictable from models and can vary by integer jumps occasionally (Blewitt, 1990). For an initial solution, the carrier phase biases can be nominally assumed to be zero (because they do not affect the light-time equation) and then estimated as real-valued parameters. If the real-valued carrier phase bias parameters are sufficiently precise, then their double differences can be resolved to their correct integer values. The solution can then be computed again by holding these integer values fixed, or more efficiently, the change to the solution can be computed using least-squares adjustment theory.

For undifferenced data processing, such as PPP, ambiguity resolution presents a conundrum. A solution to this problem lies in the stability of undifferenced phase biases. It turns out that the double difference in the estimates of the phase biases can be resolved to integer numbers of cycles (Blewitt, 1989). In the case of PPP with a single receiver (Zumberge et al., 1997), we also need estimates of biases from at least one other station with the same satellites in view. A successful method now in common use, which was proposed and demonstrated by Bertiger et al. (2010), is known as the wide-lane and phase bias (WLPB) method. The WLPB method estimates biases for the IGS network during the GPS orbit estimation procedure. These phase bias estimates are then made available to the single-receiver user along with the GPS orbit and clock parameters. Thus, the user can perform ambiguity resolution when PPP with a single receiver.

In summary, this has been a key section, in that the light eqn [6] represents the heart of GPS observable modeling and a very specific component given by eqn [14] is the source of all geophysical applications that relate to precise positioning.

3.11.2.5 Data Processing Software

Geodetic GPS data processing as implemented by research software packages can typically be generalized as a modular scheme (**Figure 7**). In this processing model, the input is raw data from GPS receivers, and the processing stops with the production of a set of station coordinates. Before discussing data processing in detail, it should be noted that the data processing does not stop with the initial production of station coordinates, but rather this is the first step toward time series analysis, velocity estimation, and kinematic analysis, all leading to dynamic analysis and geophysical interpretation. It is convenient to separate the actual processing of GPS data shown in the preceding text from the subsequent kinematic analysis, though for some geophysical applications (e.g., ocean tidal loading), this division is not correct, and geophysical parameters must be estimated directly in the solution.

Several software packages have been developed since the 1980s that are capable of delivering high-precision geodetic estimates over long baselines. These days, the processing of GPS data by these software packages is, to a large degree, automatic, or at least a 'black-box' approach is common. The black box can of course be tampered with for purposes of research into GPS methodology, but one big advantage of automation is reproducibility and consistency of results produced in an objective way.

Geodetic data processing software is a result of intensive geodetic research, mainly by universities and government research laboratories. Typical features of such software include

- orbit integration with appropriate force models;
- accurate observation model (Earth model, media delay, etc.) with rigorous treatment of celestial and terrestrial reference systems;
- reliable data editing (cycle slips and outliers);
- estimation of all coordinates, orbits, tropospheric bias, receiver clock bias, polar motion, and Earth spin rate;
- ambiguity resolution algorithms applicable to long baselines;
- estimation of reference frame transformation parameters and kinematic modeling of station positions to account for plate tectonics and coseismic displacements.

The typical quality of geodetic results from processing 24 h of data can be summarized as follows (IGS, http://igscb.jpl.nasa.gov):

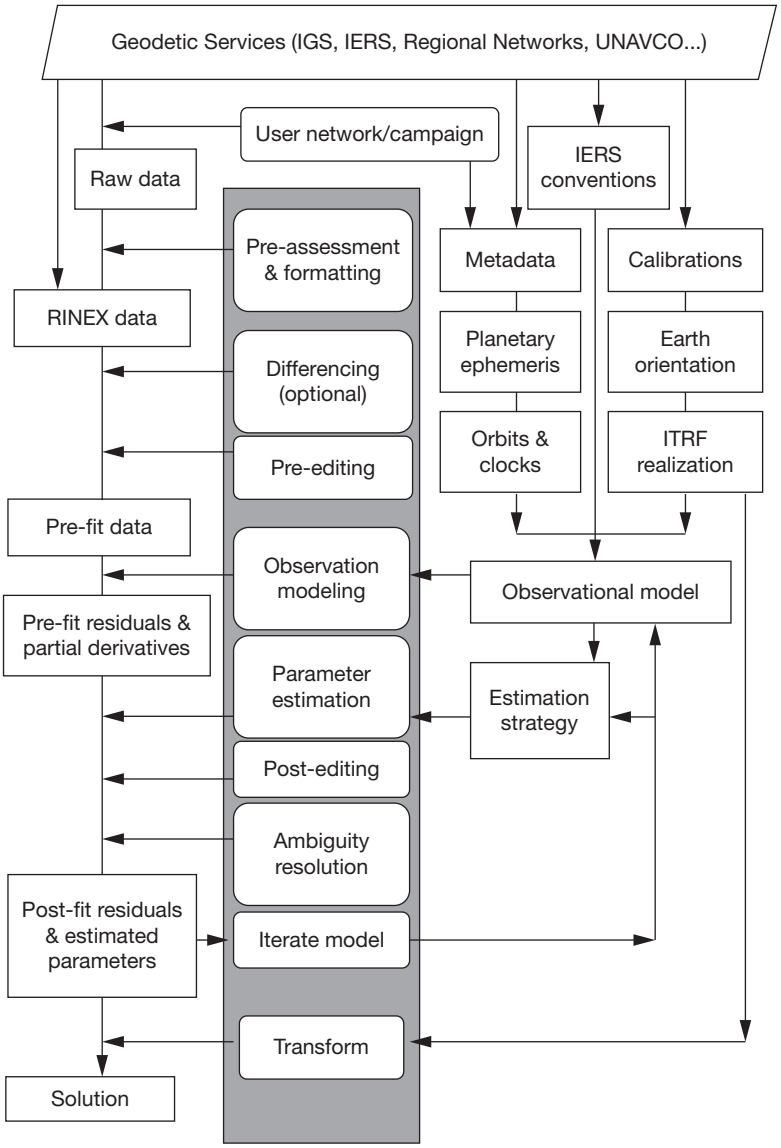


Figure 7 Generic modular scheme for geodetic GPS data processing.

- Relative positioning at the level of few parts per billion of baseline length
- Geocentric (global) positioning to ~6 mm vertical and ~3 mm horizontal in the ITRF
- Tropospheric delay estimated to ~4 mm
- GPS orbits determined to ~25 mm
- Earth pole position determined to ~2 mm
- Clock synchronization (relative bias estimation) to < 0.1 ns
- Ionospheric TEC maps to < 10 TEC units

Two features of commercial software are often conspicuously absent from more advanced packages: (1) Sometimes, double differencing is not implemented, but instead, undifferenced data are processed, and clock biases are estimated; (2) network adjustment using baseline solutions is unnecessary, since advanced packages do a rigorous, one-step, simultaneous adjustment of station coordinates directly from all available GPS observations.

Some precise software packages incorporate a Kalman filter (or an equivalent formulism) (Bierman, 1977; Herring et al., 1990; Lichten and Border, 1987). This allows for certain selected parameters to vary in time, according to a statistical ('stochastic') model. Typically, this is used for the tropospheric bias, which can vary as a random walk in time (Tralli and Lichten, 1990). A filter can also be used to estimate clock biases, where 'white noise' estimation of clock bias approaches the theoretical equivalent of double differencing.

Although many more packages have been developed, there are three ultrahigh-precision software packages, which are widely used around the world by researchers and are commonly referenced in the scientific literature:

- Bernese software, by Astronomical Institute, University of Bern, Switzerland (Rothacher et al., 1990);
- GAMIT-GLOBK software, by Massachusetts Institute of Technology, the United States (King and Bock, 2005);

- GIPSY-OASIS II software, by JPL, California Institute of Technology, the United States (Webb and Zumberge, 1993).

There are several other packages, but they tend to be limited to the institutions that wrote them. It should be noted that, unlike commercial software packages, the use of the previously mentioned software can require a considerable investment in time to understand the software and how best to use it under various circumstances. Expert training is essential.

3.11.2.6 Real-Time GPS and Accuracy Versus Latency

The most accurate GPS geodetic solutions result from a simultaneous estimation for satellite orbit and clock parameters, Earth rotation parameters, and station coordinates. One the other hand, solutions that very closely approximate those from simultaneous estimation are possible in a two-step process, where an analysis center first computes the satellite orbit and clock parameters and Earth rotation parameters and then distributes these to users who simply solve for their own station coordinates (along with the station clock, tropospheric delay, and carrier phase biases). This is basically the model used by IGS and its analysis centers to serve its users.

One consequence of the two-step model is that there is a delay for the user, who must wait until the orbit parameters become available. In the extreme case of real-time positioning, the two-step model becomes challenging to implement accurately. In this case, the orbit positions must be predicted ahead in time, and the satellite clocks must be handled in real time, either by the user differencing data or by the analysis center providing real-time estimates of the satellite clock biases.

In this section, we consider the effect of latency on accuracy and precision of point positioning for the user with one of the best software packages available. We also consider the effect of positioning interval, for example, whether the position represents a 24 h average or a new position every data epoch, which could be every 5 min, every 1 s, or even shorter intervals. Positioning at every data epoch is sometimes called 'kinematic positioning,' which is somewhat of a misnomer, as positions are directly estimated rather than kinematic parameters such as velocity (which can be subsequently inferred from the position time series). Nevertheless, we use the term 'kinematic positioning' because it has become a *de facto* standard.

The most accurate 'final' orbits and clocks are available with 12–18 day latency (some as low as 5 days). These allow the user to perform daily PPP in the ITRF to a precision of ~2 mm horizontal and ~6 mm vertical using 24 h RINEX files. For kinematic positioning, the precision degrades to ~7 mm horizontal and ~20 mm vertical, as measured by RMS scatter over a day. However, the RMS precision can be much better than this over shorter time windows, so it is possible to track very accurately the variations in position caused by seismological waves. When using the term 'precision,' therefore, one must be careful to consider the time window relevant to the problem at hand.

GPS analysis centers also produce the so-called rapid orbits with 1–2 day latency. These products are almost as good as final orbits but can differ to the extent that quality control may not be as effective at shorter latency. But as a consequence, positioning precision using rapid orbits is of similar quality to that using final orbits.

Pushing to lower latencies, GPS analysis centers also produce the so-called ultrarapid orbits with several-hour latency (some as low as 1–2 h latency). These tend to be based on the analysis of hourly RINEX files that have been concatenated over the last day or so. Ultrarapid orbits are useful for low-latency kinematic positioning problems, such as measuring the permanent displacement of stations following a large earthquake, and for rapidly inferring earthquake rupture parameters. Kinematic positioning using ultrarapid orbits is generally only slightly worse than using rapid orbits, at the 10 mm level horizontal and 30 mm level vertical. However, solutions can suffer because of several effects: (1) quality control becomes more difficult at lower latency; (2) toward the end of the ultrarapid orbit (closer to real time), the solution is based on data from the past, but not from the future, and so are degraded; (3) moreover, toward the end of the ultrarapid orbit, it becomes more difficult to detect and correct for the effect of cycle slips and to resolve integer ambiguities. As a consequence, the quality of positioning using ultrarapid orbits and clocks can be quite good but can also be sporadically poor in a rather unpredictable way.

Finally, at the extreme of real-time point positioning, it is sufficient to predict the future position of the ultrarapid orbits. However, the satellite clocks cannot be predicted accurately, and the issue of quality control becomes paramount. Moreover, the user is faced with the problem of having very poor initial estimates of carrier phase biases whenever new satellites come into view or whenever a cycle slip is detected. Carrier phase bias resets can cause systematic deviations of kinematic positions that do not reflect the actual motion of the station. Current real-time position precision is typically better than one decimeter but can be better or worse depending on the multipath and sky-view environment of the both the user's station and the stations used to determine the satellite clocks. Needless to say, improving real-time positioning is an emerging area of intense research.

3.11.3 Global and Regional Measurement of Geophysical Processes

3.11.3.1 Introduction

Geodesy is the science of the shape of the Earth, its gravitational field, and orientation in space and is therefore intrinsically connected to geophysics (Lambeck, 1988; Torge, 2001). Indeed, space geodetic techniques, such as GPS, can be used to observe the Earth and hence probe geodynamic processes on a global scale (Figure 8). GPS contributes to geophysics through comparing the observed and modeled motion of the Earth's surface. Since the observed motion of the Earth's surface will represent the sum of the various effects, it is clear that geophysics must be modeled as a whole, even when investigating a specific problem. This creates a rich area of interdisciplinary research.

As the precision and coverage of GPS stations have improved over the last two decades, the depth and breadth of GPS geodesy's application to geodynamics have increased

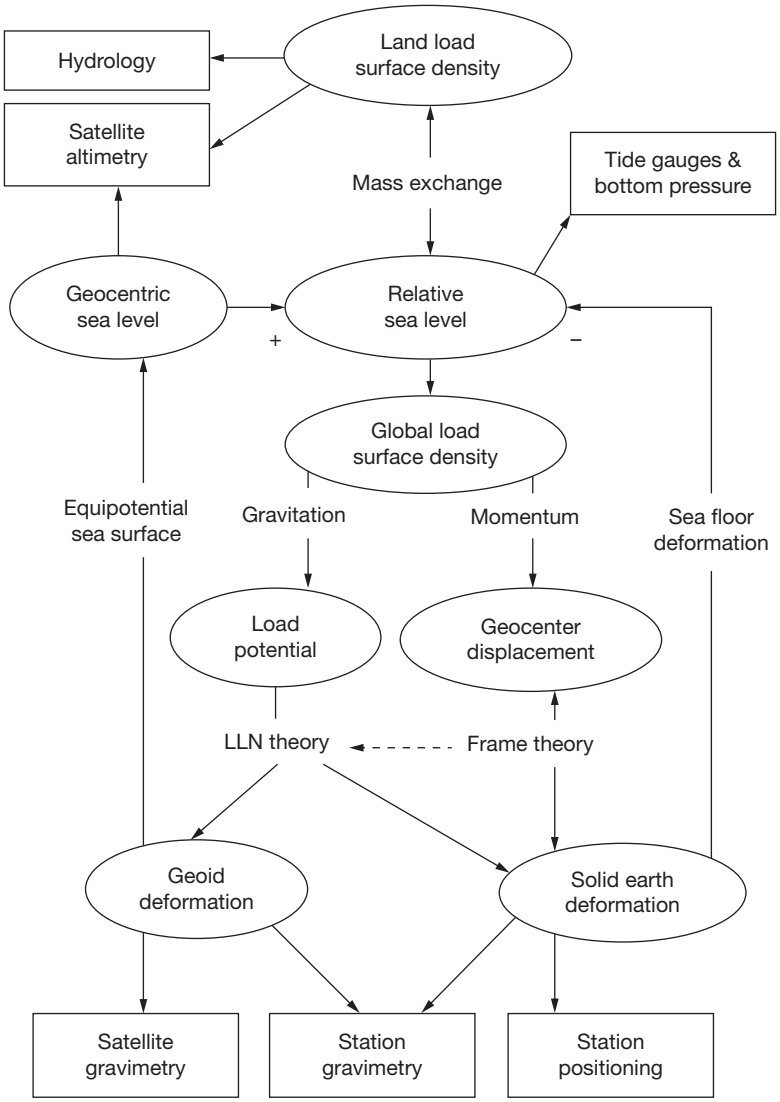


Figure 8 Schematic model of surface mass loading that incorporates self-consistency of the reference frame, loading dynamics, and passive ocean response. Closed-form inversion solutions have been demonstrated (Blewitt and Clarke, 2003; Gross et al., 2004). Note that everything is a function of time, so 'continental water' in its most general sense would include the entire past history of ice sheets responsible for postglacial rebound. (Arrows indicate the direction toward the computation of measurement models, phenomena are in round boxes, measurements are in rectangles, and physical principles label the arrows.)

correspondingly. It has now matured to the point that it is viewed as an important and often primary tool for understanding the mechanics of Earth processes.

On the other hand, geophysical models are essential to GPS geodesy; as such, models are embedded in the reference systems we use to define high-accuracy positions. For example, if the reference system did not account for the tidal deformation of the solid Earth, the coordinates of some stations could vary as much as ~10 cm in the time frame of several hours. Therefore, reference systems to enable high-accuracy geodetic positioning have developed in parallel with progress in geodynamics, which in turn depends on geodetic positioning. Thus, this interdependent relationship between geodesy and geophysics is inextricable.

Table 3 shows examples of the various geophysical processes that affect space geodetic observables and thus are subject to investigation using space geodesy. Most of the applications assume the ability to track the position of geodetic stations with sub-centimeter precision, but other possibilities include the determination of the Earth's polar motion and rate of rotation, low-degree gravitational field coefficients, and atmospheric delay in the troposphere and ionosphere. For example, global climate change not only could affect both the shape and gravitational field through mass redistribution (e.g., melting polar ice caps) but also could affect large-scale tropospheric delay.

In this section, the focus will be on providing examples of geodetic applications across the spatiotemporal spectrum, ranging from coseismic rupture and seismic waves to plate rotations.

The subsequent section will then focus on how geodesy can be used to address large-scale loading problems.

Table 3 Geophysical processes that affect geodetic observations as a function of spatial and temporal scale

Scale	*Temporal*				
Spatial	*10^{-2}–10^{3} s*	*10^{0}–10^{1} h*	*10^{0}–10^{2} day*	*10^{0}–10^{2} year*	*10^{2}–10^{6} year*
10^{0}–10^{1} km	Coseismic rupture Volcanism	Creep events Volcanism	Afterslip Poroelastic relaxation Dike injection	Viscoelastic relaxation Interseismic strain	Earthquake cycle
10^{1}–10^{2} km	M 6–7.5 seismic strain release Tropospheric moisture	Storm-surge loading Tsunami loading Tropospheric moisture	Rifting events Aquifer deformation Poroelastic relaxation Lower crustal magmatism Lake loading Snow loading	Viscoelastic relaxation Block rotation Strain partition Mountain growth Glacial loading Sedimentary loading	Fault activation and evolution Mountain range building Denudation Regional topography Sedimentary loading
10^{2}–10^{3} km	M 7.5–9 seismic strain release Traveling ionospheric disturbances Seismic waves	Coastal ocean loading	Atmospheric loading Regional hydrologic loading	Mantle–crust coupling Ice-sheet loading	Plateau rise Mountain range building Glacial cycle
10^{3}–10^{4} km	M 9+ seismic strain release Seismic waves Free oscillations	Earth tides Tidal loading	Seasonal fluid transport Ocean bottom pressure	Core–mantle coupling Climate change Solar cycle	Plate rotations Mantle flow Continental evolution

3.11.3.2 Estimation of Station Velocity

All of the following examples use a time series of discrete station positions (whether they be relative station positions or with respect to the reference frame origin). For many applications, it is convenient to first fit a 3-D station velocity to each time series. If the velocity is intended to represent the secular motion of a station but the time series spans less than ~4.5 years, then it is important to simultaneously fit an empirical seasonal signal (Blewitt and Lavallée, 2002). The simplest seasonal model would fit amplitudes for an annual sine and cosine wave. In some locations, the semiannual signal may also be important, for example,

$$\mathbf{x}_i(t_j) = \mathbf{x}_{i0} + \dot{\mathbf{x}}_i(t_j - t_0) + \mathbf{a}_{1\,i}^{C}\cos(2\pi t_j) + \mathbf{a}_{1\,i}^{S}\sin(2\pi t_j) + \mathbf{a}_{2\,i}^{C}\cos(\pi t_j) + \mathbf{a}_{2\,i}^{C}\sin(\pi t_j) + \mathbf{v}_{ij} \quad [17]$$

where $\mathbf{x}_i(t_j)$ is the observed vector position of station i at epoch t_j (in Julian years); t_0 is an arbitrary user-specified time to define the modeled epoch position $\mathbf{x}_{i0}$; $\dot{\mathbf{x}}_i$ is the station velocity (independent of the choice of t_0); the harmonic vector amplitude $\mathbf{a}_{2i}^{C}$, for example, indicates the cosine amplitude of frequency two cycles per year at station i; and $\mathbf{v}_{ij}$ represents the vector error. When using geocentric Cartesian coordinates, it is especially important to use a full 3×3 weight matrix in the inversion, because of the large difference (~factor of 3) in the magnitude of formal error in the vertical direction.

If the geophysical signals under investigation are seasonal in nature, then of course, the harmonic amplitudes are interesting in their own right, and the velocity term may be considered the 'nuisance parameter.' It should be always kept in mind that the parameter estimates will absorb the sum of all relevant geophysical processes and errors that affect the specified data set. Seasonal systematic errors are particularly difficult to quantify. In the case of simultaneous geophysical processes, it is often the case that the larger-scale processes (e.g., global-scale plate tectonics) can be characterized first and used as a calibration or as boundary conditions for a smaller-scale study (e.g., plate boundary deformation).

For some applications, it may be sufficient to study the postfit residual time series (i.e., estimates of $\mathbf{v}_{ij}$). However, caution is warranted if the form of the signal under investigation significantly is likely to correlate with the velocity or harmonic amplitude parameters. If the exact form of the signal is known (e.g., a step function in the case of a coseismic displacement field), then it is always better to augment the model in the preceding text and estimate the extra parameters simultaneously with the previously mentioned base set of parameters. On the other hand, if the exact form is not known but the signal is assumed to start with an event at a given time T, then a reasonable approach is to estimate the base parameters using only data prior to time T and then form the residuals time series for all data using this model.

Finally, it should be noted that for some inversion problems, it would be more rigorous to incorporate a stochastic model that accounts for temporal correlations in the position time series. There have been several attempts to infer the stochastic nature of errors in the time domain from spectral analysis (Mao et al., 1999; Williams, 2003). The consensus conclusion of such investigations is that GPS time series has the characteristics of flicker noise. The presence of random walk noise, which is quite damaging to the determination of station velocity, for example, is much less conclusive. The importance of these models has proven to lie largely in the realistic assignment of error bars on the estimated geophysical parameters and not so much on the actual estimates themselves. Ultimately, the accuracy of geophysical parameter estimates is better inferred by other external means, such as the smoothness of the inverted velocity field in regions where smoothness is expected from geologic considerations.

As a general rule, estimation of station velocity can be achieved with precision <1 mm year^{-1} using >2.5 years of continuous data. One measure of precision is to infer it from the smoothness of a velocity field across a network (Davis et al., 2003). In some sense, this approach gives a measure of 'accuracy,' because the results are being compared with an assumed truth (the smoothness of the velocity field). Another method, which assesses the level of systematic error, is to compare results using different software packages. For example, Hill and Blewitt (2006) compared velocities produced using the GAMIT and GIPSY software packages, where GAMIT processes double-difference data and GIPSY processes undifferenced data. Using 4 years of data from a 30-station regional GPS network, they found that the RMS difference in GPS horizontal velocity is <0.1 mm year (after accounting for a 14-parameter reference frame transformation between the two solutions). The data processing by both packages was done in a black-box fashion, with minimal user intervention. This result indicates that errors in GPS station velocities are more than likely to be dominated by biases in common to both GIPSY and GAMIT, for example, multipath error, antenna phase center mismodeling, and nonsecular Earth deformations.

3.11.3.3 Plate Tectonic Rotations

Once geodetic station velocities have been estimated (as outlined earlier), plate tectonic rotations can be estimated using the following classical kinematic model (Larson et al., 1997):

$$\dot{\mathbf{x}}_j^p = \boldsymbol{\Omega}^p \times \mathbf{x}_j \qquad [18]$$

where $\boldsymbol{\Omega}^p$ is the angular velocity (sometimes called the 'Euler vector') of a plate called 'p' associated with station j. The magnitude $\Omega^p = |\boldsymbol{\Omega}^p|$ is the 'rate of rotation' of plate p (often expressed as degrees per million years but computationally as radians per year), and the direction $\hat{\boldsymbol{\Omega}}^p = \boldsymbol{\Omega}^p / \Omega^p$ is called the 'Euler pole' (often expressed as a spherical latitude and longitude but computationally as Cartesian components, i.e., direction cosines) (Minster and Jordan, 1978). The Euler pole can be visualized as the fixed point on the Earth's surface (not generally within the plate itself) about which the plate rotates. This rotation model essentially constrains the plate to move rigidly on the Earth's surface (no radial motion). The cross product is taken between the angular velocity and station position in a geocentric reference frame; therefore, the velocity is also expressed in the geocentric reference frame. The label p on $\dot{\mathbf{x}}_j^p$ simply identifies the assumed plate (not the reference frame). This notation becomes useful later when considering the relative motion at a plate boundary.

Figure 9 shows an example of an inversion of GPS velocities for rigid plate rotations from the REVEL model (Sella et al., 2002). In this figure, only the stations so indicated were used to invert for plate rotations, on the assumption that they are located on stable plate interiors. Stations that fall within deforming plate boundaries must be treated differently, as will be explained in the following subsection.

Several points are worth noting about the classical kinematic model of plate tectonics:

- The motions are instantaneous, in the sense that the time of observation is sufficiently short that the angular velocities are assumed to be constant in time. As the equation apparently works well for paleomagnetic data over a few million

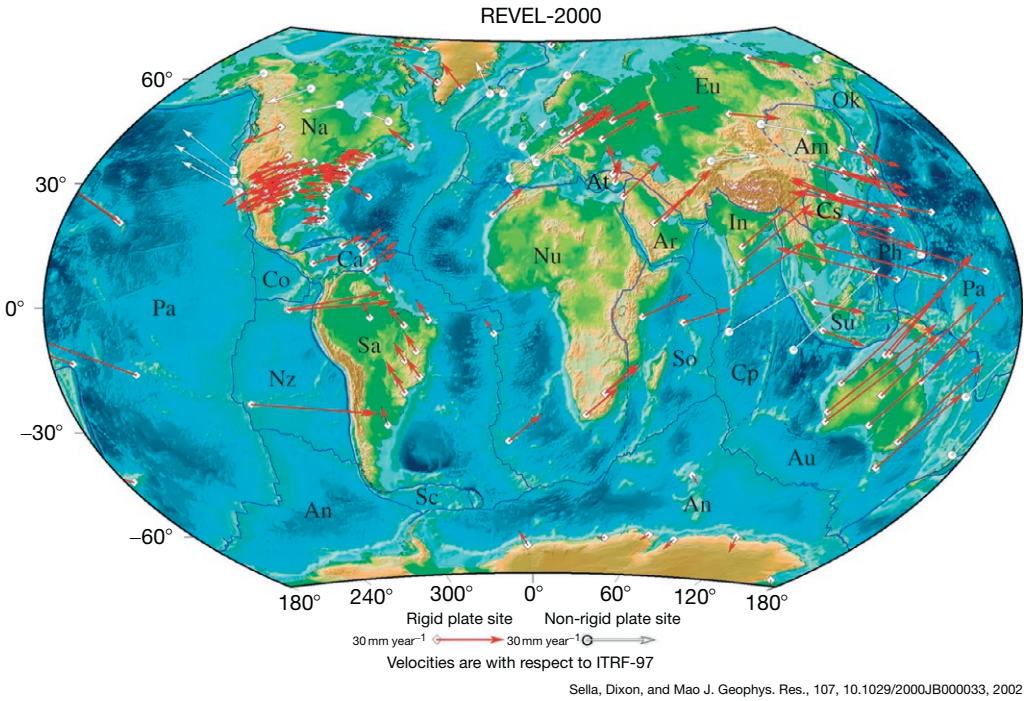


Figure 9 The REVEL-2000 plate motion model derived from GPS velocities. Reproduced from Sella G, Dixon T, and Mao A (2002) REVEL: A model for recent plate velocities from space geodesy. *Journal of Geophysical Research* **107**(B4), http://dx.doi.org/10.1029/2000JB000033.

years (DeMets et al., 1990, 1994; Minster and Jordan, 1978), such an assumption is essentially perfect for geodetic observation periods of decades. Indeed, discrepancies between angular velocities from geodesy and paleomagnetic inversions can test whether plates might have significant angular accelerations.

- Plate tectonic theory here assumes that plate motions are rigid and that the motion is a rotation about a fixed point in common to all the Earth's surface. Thus, the motions are purely horizontal on a spherical Earth.
- The assumption of plate rigidity can be tested independently of the previously mentioned model, for example, by observing changes of distance between stations supposedly on the same plate. Thus, by using geodesy (together with independent evidence), the 'stable plate interior' can be defined empirically as the domain within a plate that, to within the errors of observations, is consistent with having zero deformation. The earlier equation is therefore more properly applied to such defined 'stable plate interiors.' The relative motions between neighboring plate interiors therefore impose boundary conditions on the deformational processes that are taking place in the plate boundary region (Stein, 1993).
- Since the Earth is only approximately spherical, the earlier equation gives long systematic errors at the level of $\sim$0.2 mm year^{-1}, including in the vertical direction (with respect to the WGS 84 reference ellipsoid). Because the errors have a very long wavelength, the induced artificial strain rates are negligible ($<$0.1 nstrain year^{-1}).
- Even though vertical motions are predicted to be zero in the model, it is convenient to invert the earlier equation using Cartesian coordinates and using the full weight matrix (inverse covariance) associated with the Cartesian components of velocity. In any case, the resulting estimate of angular velocity will not be sensitive to errors in vertical velocity.
- If the true plate motions (for the part of plates exposed on the Earth's surface) are on average gravitationally horizontal (with respect to the geoid), then on average, the motion must also be horizontal with respect to the reference ellipsoid (which is defined to align with the geoid on average). Such a reference ellipsoid is necessarily centered on the center of mass of the entire Earth system, CM. Therefore, the fixed point of rotation can be taken to be CM, which is the ideal origin of ITRF. Due to the (verifiable) assumption that plate motions are constant, it is therefore important to use the long-term average CM rather than the instantaneous CM, which can move by millimeters relative to the mean Earth's surface (CF) over tidal and seasonal timescales (caused by redistribution of fluid mass).
- Any systematic error in the realization of CM will map into errors in the model for plate motions and hence errors in estimates of plate angular velocities. Significantly, this will also affect model predictions of the relative velocities of stations across plate boundaries. Considering the velocity of station j, which resides nominally on plate p (e.g., Pacific), in a reference frame corotating with plate n (e.g., North America), then this can be expressed as the following relative velocity:

$$\begin{aligned} \dot{\mathbf{x}}_j^p &= \boldsymbol{\Omega}^p \times \mathbf{x}_j \\ \dot{\mathbf{x}}_j^p - \dot{\mathbf{x}}_j^n &= \boldsymbol{\Omega}^p \times \mathbf{x}_j - \boldsymbol{\Omega}^n \times \mathbf{x}_j \\ \Delta\dot{\mathbf{x}}_j^{pn} &= \Delta\boldsymbol{\Omega}^{pn} \times \mathbf{x}_j \end{aligned} \qquad [19]$$

- Here, $\Delta\boldsymbol{\Omega}^{pq}$ is the relative angular velocity between plates p and n, and $\Delta\dot{\mathbf{x}}_j^{pn}$ is the relative velocity between plates p and n at station j, in other words, the relative velocity of station j on plate p as viewed by an observer fixed to plate n. Note that if station j actually lies in the stable interior of plate p, then $\Delta\dot{\mathbf{x}}_j^{pn}$ represents the path integral of deformation plus rotation crossing the entire plate boundary going from the stable interior of plate n to station j. Hence, systematic errors in $\Delta\boldsymbol{\Omega}^{pq}$ will negatively impact geophysical inferences on plate boundary deformation. This proves that it is important to plate tectonic applications of geodesy to realize the origin of the reference frame as the long-term center of mass of the entire Earth system. Thus, from a physical standpoint, the SLR technique is essential (to realize the origin), even if it is not the primary tool for observing relative motions between plates. It is possible to realize an appropriate origin geometrically, by assuming that, after accounting for known geophysical processes that cause vertical motion (e.g., glacial isostatic adjustment), there should be no residual vertical motion in some average sense. There are various possible ways to define such an origin, and it remains a promising topic of research to understand which types of global reference frames (in terms of their realization of the velocity reference at the origin) are most appropriate for determining plate angular velocities.
- The angular velocity parameters for any given plate are going to be best constrained by a network that maximally spans the rigid plate interior. This presents a problem if the plate is small. For very small plates (e.g., blocks in plate boundary zones), the motion can be characterized by a horizontal translation to within the sensitivity of geodetic measurements. In this case, there is a high correlation between the rate of rotation and the location of the Euler pole normal to the direction plate motion, and so the concepts of rate of rotation and Euler pole essentially lose their meaning. Nevertheless, what is important to geophysical processes is not the precision of the Euler pole and rate or rotation, but rather the precision to which relative motion is known across plate boundaries. Generally, this will be constrained very well if the geodetic network spans those boundaries.

3.11.3.4 Plate Boundary Strain Accumulation

Approximately 85% of the Earth's surface can be characterized by rigid plate tectonics. The remaining 15% can be characterized as plate boundary zones, within which the Earth's crust deforms to accommodate the relative rotation between neighboring plates (Holt et al., 2005). As these zones are responsible for generating destructive earthquakes, they are the subject of intense geodetic research. To accommodate crustal deformation, the model for rigid plate rotations can be modified to a continuum velocity field $\dot{\mathbf{x}}(\mathbf{x})$ as follows:

$$\dot{\mathbf{x}}(\mathbf{x}) = \boldsymbol{\Omega}(\mathbf{x}) \times \mathbf{x} \qquad [20]$$

where $\dot{\mathbf{x}}(\mathbf{x})$ has been parameterized in terms of a continuum angular velocity field $\boldsymbol{\Omega}(\mathbf{x})$, otherwise known as the 'rotational

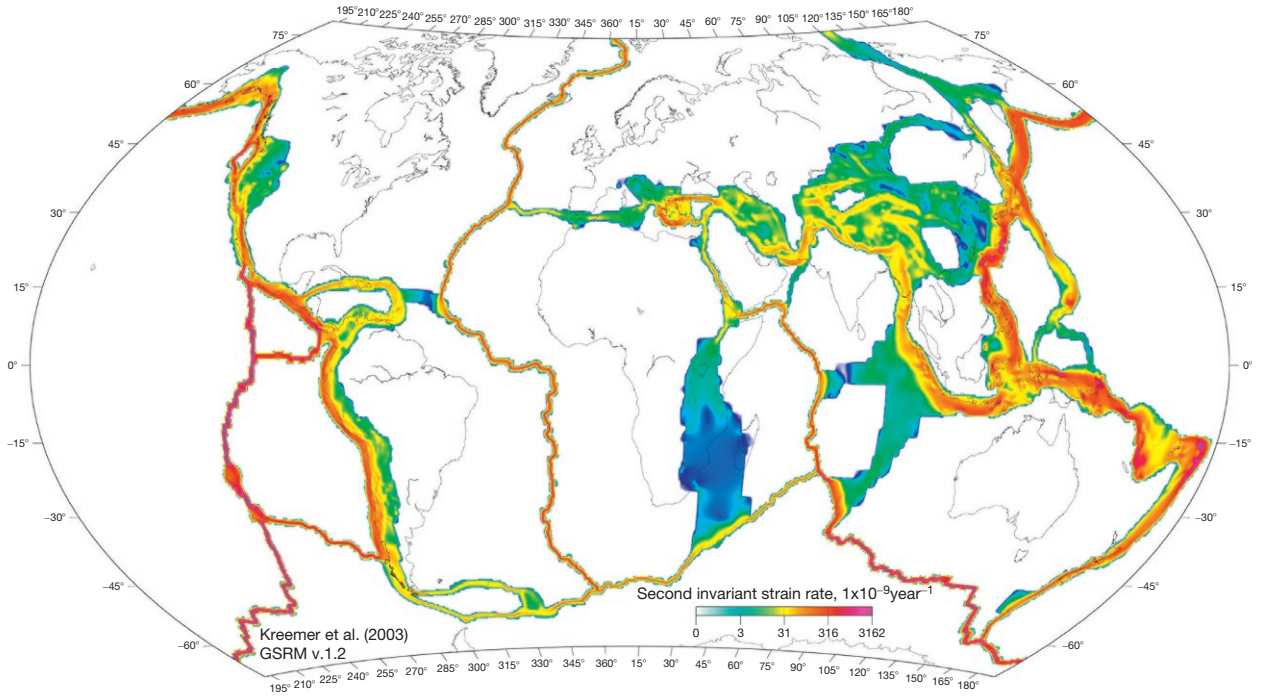


Figure 10 A global strain-rate map, derived from space geodetic data. Adapted from Kreemer C, Holt WE, and Haines AJ (2003). An integrated global model of present-day plate motions and plate boundary deformation. *Geophysical Journal International* **154**, 8–34.

vector function' (Haines and Holt, 1993). The advantage of this reparameterization is that the angular velocity field is a constant within a stable plate interior, unlike the velocity field, which appears as a rotation, depending on the defined reference frame. If a region can be defined a priori as being on a stable plate interior, then $\boldsymbol{\Omega}(\mathbf{x})$ can be constrained as a constant parameter in the model: $\boldsymbol{\Omega}(\mathbf{x}) = \boldsymbol{\Omega}^p$, and in these regions, the formula reduces to the plate rotation model. Otherwise, spatial gradients of $\boldsymbol{\Omega}(\mathbf{x})$ correspond to deformation rates. Specifically, the three horizontal components of the deformational (symmetrical and nonrotating) strain-rate tensor on a sphere can be written as

$$\begin{aligned}\dot{\boldsymbol{\varepsilon}}_{\varphi\varphi} &= \frac{\hat{\boldsymbol{\Theta}}}{\cos\theta}\bullet\frac{\partial\boldsymbol{\Omega}}{\partial\varphi}\\ \dot{\boldsymbol{\varepsilon}}_{\theta\theta} &= -\hat{\boldsymbol{\Phi}}\bullet\frac{\partial\boldsymbol{\Omega}}{\partial\theta}\\ \dot{\boldsymbol{\varepsilon}}_{\varphi\theta} &= \frac{1}{2}\left(\hat{\boldsymbol{\Theta}}\bullet\frac{\partial\Omega}{\partial\theta} - \frac{\hat{\boldsymbol{\Phi}}}{\cos\theta}\bullet\frac{\partial\boldsymbol{\Omega}}{\partial\varphi}\right)\end{aligned} \qquad [21]$$

where $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\Phi}}$ are unit vectors that point in the north and east directions, respectively. The contribution of vertical velocity to horizontal strain rates is neglected, because this is $<2\%$ for even rapid uplift rates of 10 mm year^{-1}. Similarly, the vertical component of the rotation rate (the symmetrical strain-rate tensor component) is

$$w = \frac{1}{2}\left(\hat{\boldsymbol{\Theta}}\bullet\frac{\partial\boldsymbol{\Omega}}{\partial\theta} + \frac{\hat{\boldsymbol{\Phi}}}{\cos\theta}\bullet\frac{\partial\boldsymbol{\Omega}}{\partial\varphi}\right) \qquad [22]$$

The Global Strain Rate Map (GSRM) Project has implemented this approach to invert GPS station velocities for a global map of strain (Holt et al., 2005; Kreemer et al., 2000, 2003). The GSRM website (http://www.world-strain-map.org) is housed and maintained at UNAVCO facility in Boulder, Colorado. The GSRM website has an introduction page where one can access information on the methodology used, the data and references, model results, and acknowledgments. A sample of a global strain-rate map is presented in **Figure 10**. Areas with no color (white) are constrained a priori to be stable (zero strain, corresponding to rigid plate rotation, as in the classical plate tectonic model).

Haines (1982) showed that if the spatial distribution of strain rates is everywhere defined, then the full velocity gradient tensor is uniquely defined. Given that geodetic stations only sample the continuum velocity field at a discrete set of locations, additional constraints are required to invert the equations. In early work, the rotation vector function was expanded as polynomials (Holt and Haines, 1993; Holt et al., 1991; Jackson et al., 1992), but in all later work, the bicubic Bessel interpolation on a curvilinear grid has been used for the Aegean Sea (Jackson et al., 1994), Asia (Holt et al., 1995), Iran (Jackson et al., 1995), Japan (Shen-Tu et al., 1995), the Indian Ocean (Tinnon et al., 1995), the western United States (Shen-Tu et al., 1999), New Zealand (Holt and Haines, 1995), and the Tonga subduction zone (Holt, 1995).

The art of designing appropriate constraint methods is a fertile area of research. In general, the constraints should be data-driven where there are data but should be averse to generating artifacts in sparsely sampled regions. Ideally, the constraints should adapt the spatial resolution to the extreme nonhomogeneity that is the case for today's global network of continuous GPS stations. However, it should be kept in mind that no matter what the constraints, they will generally smooth the observed velocity field to some extent and will

generally generate anomalous spatially distributed artifacts around stations with velocity errors. Such is the nature of underdetermined inversion problems. The key to successful geophysical interpretation is to not overanalyze the results and to use only information at wavelengths longer than a spatial resolution appropriate to the expected errors. One exception to this is the case where anomalous motion of a station is truly geophysical and related to an interesting localized process. Strain-rate mapping can be used to help identify such candidates for further investigation.

One method of imposing constraints is using independent strain-rate inferences from earthquake moment tensors and geologic fault slip data, through Kostrov's relation (Kostrov, 1974). Observed average seismic strain rates for any grid area can be obtained by summing moment tensors in the volume described by the product of the grid area and the assumed seismogenic thickness:

$$\dot{\varepsilon}_{ij} = \frac{1}{2\mu VT}\sum_{k=1}^{N} M_0 m_{ij} \qquad [23]$$

where N is the number of events in the grid area, μ is the shear modulus, V the cell volume, T is the time period of the earthquake record, M_0 is the seismic moment, and m_{ij} is the unit moment tensor. Similarly, average horizontal strain-rate components from Quaternary fault slip data can be obtained by a variant of Kostrov's summation (Kostrov, 1974) over N fault segments k within a grid area A:

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\sum_{k=1}^{N} \frac{L_k \dot{u}_k}{A \sin \delta_k} m_{ij}^k \qquad [24]$$

where m_{ij}^k is the unit moment tensor defined by the fault orientation and unit slip vector and the fault segment has length L_k, dip angle δ_k, and slip rate $\dot{u}_k$.

In this combined (geodetic + seismic + geologic) scheme, an objective minimization function can then be defined that accommodates all three data types (e.g., Kreemer et al., 2000). Typically, geodetic data are given a strong weight in such schemes because, unlike the case for geodetic data, it is not clear to what extent a limited sample of earthquake moment tensors or Quaternary geologic data represent strain rates today. Whereas this approach can be applied to produce a combined (geodetic plus seismic) solution for strain-rate mapping, an alternative approach is to produce an independent empirical geodetic solution from which to compare other geophysical data types.

A different approach to strain mapping is based on the concept of 2-D tomography. The idea is that the relative velocity between distant geodetic stations must equal the path integral of strain no matter what the path. Therefore, faults can be assigned slip rates and block domains can be assigned rotations such that path integrals that cross these structures agree with the geodetic data. This approach requires the user to construct a variety of path integrals that will ensure a well-conditioned solution. Although this introduces an additional level of nonuniqueness (due to the user's choices of path integral), the resulting strain-rate maps are insensitive to the choices made so long as their choice overdetermines the problem. Implicitly, the rotation function approach also ensures that the path integral agrees in a least-squares sense with relative velocities between stations, and so both methods lead to similar solutions, assuming the a priori constraints (from nongeodetic evidence) are approximately equivalent.

The applications of strain-rate mapping in plate boundary zones are numerous, ranging from understanding mantle-scale processes to identifying areas of enhanced seismic hazard. The general pattern of the style of strain can point to the larger-scale picture of the driving dynamics. Deformations can be understood as the sum of dilatational strain (increase in surface area) and shear strain (distortion of shape). Regions of strain that are predominantly represented by shear relate to strike-slip faulting, which typically accommodates strain across transform boundaries, such as the North America–Pacific plate boundary. Positive dilatational strain is associated with zones of extension, which can be driven by a combination of gravitational collapse and boundary conditions on a region, as is the case of the Great Basin in western North America. Gravitational collapse is a predominant factor in the Himalayas and the broader zone of deformation in Southeast Asia. Negative dilatational strain is of course associated with convergent plate boundaries. Whereas these examples are largely related to mantle-scale processes, on the smaller-scale, combinations of all styles of strains can arise from inhomogeneities in the crust. For example, kinks in a strike-slip fault can create either a compressional fold or pull-apart basins.

Clearly, all these processes can be and have been studied with nongeodetic techniques and geodesy should be considered as just one tool that can be brought to bear. Broadly speaking, what geodesy brings to the table are the following two basic advantages:

- Geodesy can provide a seamless, consistent map of strain rates spanning a broad range of distance scales, ranging from seismogenic thickness (~15 km) to the global scale (~10 000 km). As such,
  - o geodesy can provide a spatial framework within which other types of geophysical evidence can be better interpreted,
  - o geodesy can indicate to what extent strain can be attributed to broader mantle-scale processes versus more localized crustal-scale structures,
  - o a geodetic map of strain rate can provide boundary conditions for a study area within which more detailed fieldwork can be preformed and understood in the broader context.
- Geodesy can provide a seamless, consistent characterization of changes in strain- rates over the timescales of seconds to decades. As such,
  - o geodesy clearly represents what is happening today (differences with other techniques may point to temporal evolution in recent geologic time),
  - o geodesy is an appropriate tool to study all phases of the earthquake cycle (the topic of the next section), ranging from coseismic rupture, through postseismic relaxation, to steady-state interseismic strain accumulation.

3.11.3.5 The Earthquake Cycle

As pointed out in the previous section, geodesy can be used to investigate motions of the Earth's surface on timescales of

seconds to decades and so is an appropriate tool to study all phases of the earthquake cycle (Hammond, 2005).

Figure 11 schematically illustrates the expected characteristics of geodetic position time series as a function of time and distance from a fault through the earthquake cycle. In this specific example, the fault is strike-slip with two stations on either side of the fault located at equal distance normal to the fault strike (Figure 12). 'Displacement' is defined as the relative position of the two stations in the direction parallel to the strike of the fault. Each curve represents a different distance from the fault (i.e., half the distance between stations). For purposes of illustration, the plot is not to scale. The plot shows the time of characteristic earthquakes EQ1 and EQ2, which are of the largest magnitude that can typically occur on this particular fault, such that smaller earthquakes produce displacements so small that they can be neglected for purposes of illustrating the earthquake cycle. Thus, EQ1 and EQ2 represent the start and end of an earthquake cycle. The size of these earthquakes ($M_W \sim 7$) is sufficient that they rupture from seismogenic depth through to the surface, with coseismic slip approximately constant with depth.

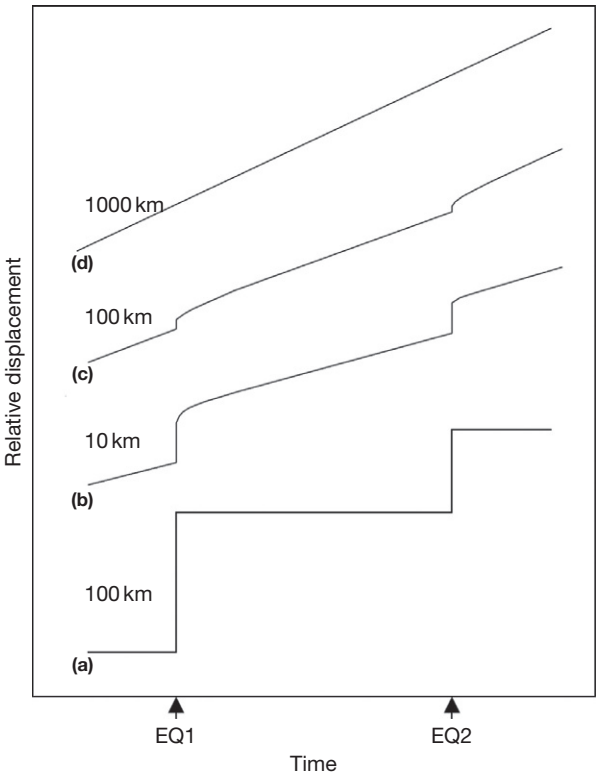


Figure 11 Schematic illustration of the effect of the earthquake cycle on geodetic station positions (see text for explanation).

Case (a) at distance 100 m shows a displacement equal to the coseismic slip on the rupture plane. In between earthquakes, the distance between stations is so small that no deformation is detected. Hence, case (a) is effectively equivalent to a geologic determination of coseismic fault slip. In the opposite extreme, case (d) at 1000 km from the fault in the far field shows no detectable coseismic displacement. (This assumes naively that this is only active fault in the region of this scale.) This displacement represents the far-field driving force transmitted through the crust that ultimately causes the earthquakes. In a sense, the earthquake represents the crust 'catching up' to where it would be if the fault were continuously sliding as a frictionless plane. Thus, case (d) shows the same average displacement per year as would a regression to curve (a) over a sufficient number of earthquake cycles. Thus, case (d) also represents (1) the slip rate at depth below the locked (seismogenic) portion of the crust, assuming the crust behaves perfectly elastically, and (2) the mean slip rate inferred by geologic observations of recent Quaternary earthquakes over several earthquake cycles, assuming that the activity of this fault is in steady-state equilibrium and is not evolving in time.

Case (b) represents the strain accumulation and release where strain rates are highest in the near field of the fault. In this case, the coseismic displacement is slightly damped due to the coseismic rupture being of finite depth. On the other hand, the time series captures subsequent near-field postseismic effects following each earthquake (Pollitz, 1997, 2003). These processes include afterslip (creep) caused by a velocity-hardening rheology and poroelastic relaxation in response to coseismic change in pore pressure. Significant aftershock might

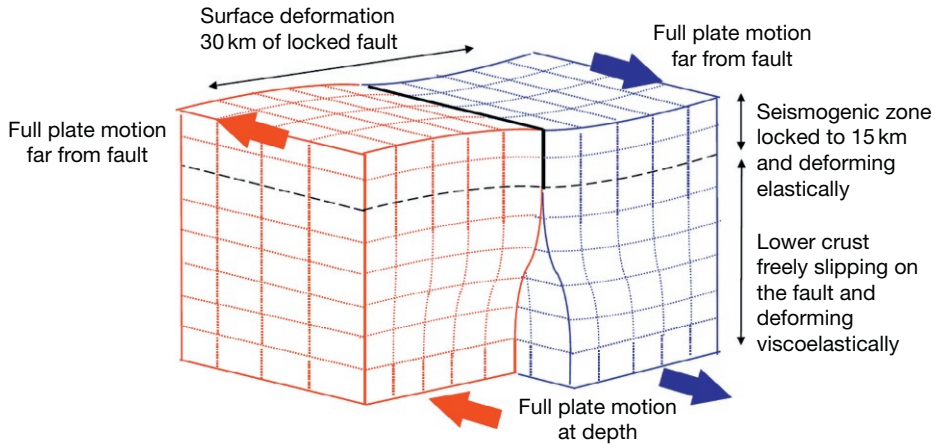


Figure 12 Deformation around a transform plate boundary, such as the San Andreas Fault.

in some cases be a contributing factor. These processes affect GPS position time series in the days to months following the earthquake (Kreemer et al., 2006a).

Over periods of years to decades, the strain for case (b) is slowly released in the viscoelastic layers beneath the crust, which will affect the time series. This occurs because at the time of the earthquake, the crust displaces everywhere and stresses the layers beneath. These layers react instantly as an elastic medium, but as time increases, they start to flow viscously. As the time series ages toward the second earthquake EQ2, most of the viscoelastic response has decayed, and the time series becomes flat. This represents the phase of interseismic strain accumulation. However, the slope of this part of the curve is significantly lessened owing to transient postseismic processes earlier in the earthquake cycle. Thus, near-field measurements of strain alone can significantly underestimate the seismic hazard unless this is taken into account. This dampening phenomenon also makes it more difficult to pinpoint the location of active faults capable of generating large earthquakes.

Stations in case (c) are still sufficiently close that coseismic displacements can be detected but are far enough from the rupture that near-field relaxation does not contribute to the time series. On the other hand, the crust in this intermediate field responds significantly to deep viscoelastic relaxation at the base of the crust and beyond, into the upper mantle. Precisely how the pattern of deformations looks in the years after a large earthquake depends on the relative effective viscosity of these various layers (Hetland and Hager, 2006). Thus, GPS networks with stations at various distances about a recently ruptured fault can be used to probe rheology versus depth. Note that in going from EQ1 to EQ2, the relative velocity between the pair of stations decreases in time. Thus, the strain rate can depend considerably on the phase of the earthquake cycle (Dixon et al., 2003). Thus, in regions of low strain where there are many faults that rupture infrequently (such as the Great Basin, western United States), it is not uncommon to observe strain rates that are almost entirely transient in nature and can exceed interseismic strain by an order of magnitude (Gourmelen and Amelung, 2005; Hetland and Hager, 2003). This makes it more difficult to interpret strain-rate maps in terms of seismic hazard, except in those cases where the strain rates are so large that logically, they must be dominated by interseismic strain accumulation on dominant faults (Kreemer et al., 2006b). Such analysis is on the leading edge of research and requires careful modeling of the earthquake cycle in any given region of interest.

3.11.3.6 Surface Mass Loading

The Earth's time-variable geometric shape, gravitational field, and rotation in space are all connected by the Earth's dynamic response to the redistribution of near-surface mass, including mass in the ocean, continental water, ice sheets, and atmosphere. As a consequence, measurements of the Earth's geometric shape from GPS can be used to infer surface mass redistribution and therefore predict changes to the gravitational field and the Earth's rotation. Thus, GPS measurements of the Earth's shape can be independently checked by comparison with time-variable gravity as measured in space by geodetic satellites or residual measurements of the Earth's rotation (i.e., change in the Earth's rotation driven by change in moment of inertia).

Surface loading therefore represents a unifying theme in geodesy, connecting various types of geodetic measurement and geodynamic models (**Figure 8**). With an assumed structure and rheology of the Earth, it becomes possible to estimate surface mass redistribution from the changing shape of the Earth (Plag et al., 1996). Conversely, with a known source of mass redistribution (e.g., inferred by gravity measurements), it should be possible to invert the measured shape of the Earth to solve for the Earth's structure (and rheology, if we include the time-variable response). That is, the ratio of the Earth's gravitational response to geometric response can be used to infer the Earth's structure and rheology.

Loading models have traditionally used Green's functions, as derived by Farrell (1972) and applied in various geodetic investigations (e.g., Van Dam et al., 1994). Green's function approach is fundamentally based on load Love number theory, in which the Earth's deformation response is a function of the spherical harmonic components of the incremental gravitational potential created by the surface load. To study the interaction between loading dynamics and the terrestrial reference frame, it is convenient to use the spherical harmonic approach (Grafarend et al., 1997; Lambeck, 1988; Mitrovica et al., 1994) (therefore, the conclusions must also apply to the use of Green's functions).

The following 'standard model' is based on a spherically symmetrical, radially layered, elastic Earth statically loaded by a thin shell on the Earth's surface. Farrell (1972) used such a model to derive Green's functions that are now prevalent in atmospheric and hydrologic loading models (van Dam et al., 2001). The preliminary reference Earth model PREM (Dziewonski and Anderson, 1981) yields load Love numbers almost identical to Farrell's (Grafarend et al., 1997; Lambeck, 1988).

It is analytically convenient to decompose the Earth system as a spherical solid Earth of radius R_E, plus surface mass that is free to redistribute in a thin surface layer ($<<R_E$) of surface density $\sigma(\Omega)$, which is a function of geographic position Ω (latitude φ and longitude λ). Let us express the total redistributed load as a spherical harmonic expansion:

$$\sigma(\Omega) = \sum_{n=1}^{\infty} \sum_{m=0}^{n} \sum_{\Phi=C}^{S} \sigma_{nm}^{\Phi} Y_{nm}^{\Phi}(\Omega) \qquad [25]$$

where $Y_{nm}^{\Phi}(\Omega)$ are defined in terms of associated Legendre polynomials: $Y_{nm}^{C} = P_{nm}(\sin\phi)\cos m\lambda$ and $Y_{nm}^{S} = P_{nm}(\sin\phi)\sin m\lambda$.

The summation begins at degree $n=1$ assuming that mass is conserved in the Earth system. It is this initial degree-one term that relates to the origin of the reference frame. It can be shown (Bomford, 1971) that, for a rigid Earth, such a thin-shell model produces the following incremental gravitational potential at the Earth's surface, which we call the 'load potential':

$$\begin{aligned} V(\Omega) &= \sum_{n} V_n(\Omega) \\ &= \frac{4\pi R_E{}^3 g}{M_E} \sum_{n} \sum_{m} \sum_{\Phi} \frac{\sigma_{nm}^{\Phi} Y_{nm}^{\Phi}(\Omega)}{(2n+1)} \end{aligned} \qquad [26]$$

where g is acceleration due to gravity at the Earth's surface and M_E is the mass of the Earth. This load potential results in a

displacement of the geoid called the 'equilibrium tide.' As shall be addressed later, the load deforms the solid Earth and, in doing so, creates an additional potential.

According to load Love number theory, solutions for surface displacements $\Delta s_h(\Omega)$ in the local height direction and $\Delta s_l(\Omega)$ in any lateral direction specified by unit vector $\hat{\mathbf{l}}(\Omega)$ are given by (Lambeck, 1988)

$$\begin{aligned}\Delta s_h(\Omega) &= \sum_n h'_n V_n(\Omega)/g \\ \Delta s_l(\Omega) &= \sum_n l'_n \hat{\mathbf{l}}(\Omega).\nabla V_n(\Omega)/g\end{aligned} \quad [27]$$

and the additional potential caused by the resulting deformation is

$$\Delta V(\Omega) = \sum_n k'_n V_n(\Omega) \quad [28]$$

where h'_n, l'_n, and k'_n are degree-n load Love numbers, with the prime distinguishing Love numbers used in loading theory from those used in tidal theory. The surface gradient operator is defined as $\nabla = \hat{\boldsymbol{\varphi}}\partial_\varphi + \hat{\boldsymbol{\lambda}}(1/\cos\varphi)\partial_\lambda$, where $\hat{\phi}$ and $\hat{\boldsymbol{\lambda}}$ are unit vectors pointing northward and eastward, respectively.

The net loading potential (load plus additional potential) relative to Eulerian observer (the 'space potential' as observed on a geocentric reference surface) is

$$\begin{aligned}U(\Omega) &= V(\Omega) + \Delta V(\Omega) \\ &= \sum_n \left(1 + k'_n\right) V_n(\Omega)\end{aligned} \quad [29]$$

The net loading potential relative to Lagrangian observer (the 'body potential' as observed on the deforming Earth's surface) must also account for the lowering of the Earth's surface due to loading. From eqns [5] and [3], the body potential is

$$\begin{aligned}U'(\Omega) &= U(\Omega) - g\Delta s_h(\Omega) \\ &= \sum_n \left(1 + k'_n - h'_n\right) V_n(\Omega)\end{aligned} \quad [30]$$

Therefore, the 'space' and 'body' combinations of load Love number, $(1+k'_n)$ and $(1+k'_n-h'_n)$, are relevant to computing gravity acting on Earth-orbiting satellites and Earth-fixed instruments, respectively.

Solutions for surface deformations of the thin-shell loading model are found by substituting [2] into [3] and [5]:

$$\begin{aligned}\Delta s_h(\Omega) &= \frac{4\pi {R_E}^3}{M_E} \sum_n \sum_m \sum_\Phi \frac{h'_n}{2n+1} \sigma^\Phi_{nm} Y^\Phi_{nm}(\Omega) \\ \Delta s_l(\Omega)^l &= \frac{4\pi {R_E}^3}{M_E} \sum_n \sum_m \sum_\Phi \frac{l'_n}{2n+1} \sigma^\Phi_{nm} \hat{\mathbf{l}}.\nabla Y^\Phi_{nm}(\Omega) \\ U(\Omega) &= \frac{4\pi {R_E}^3}{M_E} \sum_n \sum_m \sum_\Phi \frac{1+k'_n}{2n+1} \sigma^\Phi_{nm} Y^\Phi_{nm}(\Omega)\end{aligned} \quad [31]$$

Thus, GPS data on station coordinate variations around the globe can be used to invert eqn [7] for the surface mass coefficients (up to some degree and order n) and hence the surface mass field by substitution into eqn [1] (Blewitt and Clarke, 2003). Truncation of the expansion is of course necessary due to the discrete and finite coverage of GPS data, especially considering the sparsity of data in certain areas such as over the ocean. This implies that the surface mass field will be smoothed. Nevertheless, the long-wavelength information from geodesy is in principle useful to constrain the continental-scale integral of basin-scale hydrologic models.

While it is in widespread use, the previously mentioned standard loading model might be improved by incorporating the Earth's ellipticity (Wahr, 1981), mantle heterogeneity (Dziewonski et al., 1977; Plag et al., 1996; Su et al., 1994; van Dam et al., 1994), and Maxwell rheology (Lambeck, 1988; Mitrovica et al., 1994; Peltier, 1974). There is no consensus model to replace PREM yet; however, the general approach to reference frame considerations described here would be applicable to improved models.

To date, surface mass loading has primarily been investigated by gravimetric methods (e.g., GRACE and SLR, see elsewhere in this volume), and the application of geometric measurements from GPS is still in its infancy. The most promising application of GPS in this respect is to the lower-degree harmonic components of the global surface mass field, to which satellite missions such as GRACE are least sensitive.

References

Altamimi Z, Angermann D, Argus D, et al. (2001) The terrestrial reference frame and the dynamic Earth. *EOS. Transactions of the American Geophysical Union* 82: 273–279.

Altamimi Z, Sillard P, and Boucher C (2002) ITRF2000: A new release of the International Terrestrial Reference Frame for Earth science application. *Journal of Geophysical Research* 107(B10): 2214. http://dx.doi.org/10.1029/2001JB000561.

Argus DF, Peltier WR, and Watkins MM (1999) Glacial isostatic adjustment observed using very long baseline interferometry and satellite laser ranging geodesy. *Journal of Geophysical Research* 104(B12): 29077–29093.

Ashby N (2004) The Sagnac effect in the Global Positioning System. In: Rizzi G and Ruggiero ML (eds.) *Relativity and Rotating Frames (Fundamental Theories of Physics)*, vol. 135, pp. 11–28. Dordrecht: Springer.

Ashby N and Allan D (1984) Coordinate time on and near the Earth. *Physical Review Letters* 53(19): 1858.

Bar-Sever Y (1996) A new model for GPS yaw attitude. *Journal of Geodesy* 70: 714–723.

Bar-Sever YE, Kroger PM, and Borjesson JA (1998) Estimating horizontal gradients of tropospheric path delay with a single GPS receiver. *Journal of Geophysical Research* 103: 5019–5035.

Bennett RA, Wernicke BP, and Davis JL (1998) Continuous GPS measurements of contemporary deformation across the northern Basin and Range province. *Geophysical Research Letters* 25: 563–566.

Bennett RA, Wernicke BP, Niemi NA, Friedrich AM, and Davis JL (2003) Contemporary strain rates in the northern Basin and Range province from GPS data. *Tectonics* 22(2): 1008. http://dx.doi.org/10.1029/2001TC001355.

Bertiger WI, Desai SD, Haines B, et al. (2010) Single receiver phase ambiguity resolution with GPS data. *Journal of Geophysical Research* 84(5): 327–337.

Beutler G, Brockmann E, Gurtner W, Hugentobler U, Mervart L, and Rothacher M (1994) Extended orbit modeling techniques at the CODE Processing Center of the International GPS Service for Geodynamics (IGS): Theory and initial results. *Manuscripta Geodaetica* 19: 367–386.

Beutler G, Drewes H, and Verdun A (2004) The new structure of the International Association of Geodesy (IAG) viewed from the perspective of history. *Journal of Geodesy* 77: 560–575.

Beutler G, Gurtner W, Bauersima I, and Langley R (1985) Modeling and estimating the orbits of GPS satellites. In: *Proceedings of the 1st International Symposium on Precise Positioning with the Global Positioning System* Rockville, MD: U.S. Dept. of Commerce.

Beutler G, Mueller II, and Neilan RE (1994) The International GPS Service for Geodynamics: Development and start of official service on January 1, 1994. *Bulletin Geodesique* 68: 39–70.

Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, and Ware RH (1992) GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System. *Journal of Geophysical Research* 97(D14): 15787–15801.

Bianco G, Devoti R, Fermi M, Luceri V, Rutigliano P, and Sciarretta C (1997) Estimation of low degree geopotential coefficients using SLR data. *Planetary and Space Science* 46: 1633–1638.

Bierman G (1977) *Factorization Methods for Discrete Sequential Estimation*. New York, NY: Academic Press.

Bilham R (1991) Earthquakes and sea level: Space and terrestrial metrology on changing planet. *Reviews of Geophysics* 29: 1–29.

Blewitt G (1989) Carrier phase ambiguity resolution for the Global Positioning System applied to geodetic baselines up to 2000 km. *Journal of Geophysical Research* 94(B8): 10187–10283.

Blewitt G (1990) An automatic editing algorithm for GPS data. *Geophysical Research Letters* 17(3): 199–202.

Blewitt G (1993) Advances in Global Positioning System technology for geodynamics investigations. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Technology*, pp. 195–213. Washington, DC: American Geophysical Union.

Blewitt G and Clarke P (2003) Inversion of Earth's changing shape to weigh sea level in static equilibrium with surface mass redistribution. *Journal of Geophysical Research* 108(B6): 2311. http://dx.doi.org/10.1029/2002JB002290.

Blewitt G, Kreemer C, Hammond WC, Plag H-P, Stein S, and Okal E (2006) Rapid determination of earthquake magnitude using GPS for tsunami warning systems. *Geophysical Research Letters* 33: L11309. http://dx.doi.org/10.1029/2006GL026145.

Blewitt G and Lavallée D (2002) Effect of annual signals on geodetic velocity. *Journal of Geophysical Research* 107(B7). http://dx.doi.org/10.1029/2001JB000570.

Blewitt G and Lichten SM (1992) Carrier phase ambiguity resolution up to 12000 km: Results from the GIG'91 experiment. In: Proceedings of the 6th International Symposium on Satellite Positioning, Columbus: Ohio State University.

Bock Y, Gourevitch SA, Counselman CC, King RW, and Abbot RI (1986) Interferometric analysis of GPS phase observations. *Manuscripta Geodaetica* 11: 282–288.

Bock Y, Nikolaidis RM, de Jonge PJ, and Bevis M (2000) Instantaneous geodetic positioning at medium distances with the Global Positioning System. *Journal of Geophysical Research* 105(B12): 28223–28253.

Bock Y, Prawirodirdjo L, and Melbourne TI (2004) Detection of arbitrarily dynamic ground motions with a dense high-rate GPS network. *Geophysical Research Letters* 31(B10): L06604. http://dx.doi.org/10.1029/2003GL019150.

Boehm J, Niell A, Tregoning P, and Schuh H (2006) Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data. *Geophysical Research Letters* 33: L07304. http://dx.doi.org/10.1029/2005/GL025546.

Bomford G (1971) *Geodesy*, third rev. ed Oxford, UK: Oxford University Press. ISBN 978-0198519195.

Bosworth JM, Coates RJ, and Fischetti TL (1993) The Development of NASA's Crustal Dynamics Project. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Technology*, pp. 1–20. Washington, DC: American Geophysical Union.

Boucher C and Altamimi Z (1993) Development of a conventional terrestrial reference frame. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*, pp. 89–98. Washington, DC: American Geophysical Union.

Chao BF, O'Connor WP, Chang ATC, Hall DK, and Foster JL (1987) Snow load effect on the Earth's rotation and gravitational field, 1979–1985. *Journal of Geophysical Research* 92: 9415–9422.

Chen G and Herring TA (1997) Effects of atmospheric azimuthal asymmetry on the analysis of space geodetic data. *Journal of Geophysical Research* 102: 20489–20502.

Chen JL, Wilson CR, Eanes RJ, and Nerem RS (1999) Geophysical interpretation of observed geocenter variations. *Journal of Geophysical Research* 104(B2): 2683–2690.

Cheng MK and Tapley BD (1999) Seasonal variations in low degree zonal harmonics of the Earth's gravity field from satellite laser ranging observations. *Journal of Geophysical Research* 104(B2): 2667–2681, February 10.

Cheng M and Tapley BD (2004) Variations in the Earth's oblateness during the past 28 years. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2004JB003028.

Choi K, Bilich A, Larson K, and Axelrad P (2004) Modified sidereal filtering: Implications for high-rate GPS positioning. *Geophysical Research Letters* 31(22), L22608.

Ciufolini I and Pavlis EC (2004) A confirmation of the general relativistic prediction of the Lense-Thirring effect. *Nature* 431: 958–960.

Counselman CC, Hinteregger HF, and Shapiro II (1972) Astronomical applications of differential interferometry. *Science* 178: 607–608.

Counselman CC and Shapiro II (1979) Miniature interferometer terminals for Earth surveying. *Bulletin Geodesique* 53(2): 139–163.

Davis JL, Bennett RA, and Wernicke BP (2003) Assessment of GPS velocity accuracy for the Basin and Range Geodetic Network (BARGEN). *Geophysical Research Letters* 30(7): 1411. http://dx.doi.org/10.1029/2003GL016961.

Davis JL, Elgered G, Niell AE, and Kuehn CE (1993) Ground-based measurement of gradients in the "wet" radio refractivity of air. *Radio Science* 28: 1003–1018.

Davis JL, Herring TA, Shapiro II, Rogers AEE, and Elgered G (1985) Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio Science* 20: 1593–1607.

Davis JL, Murray MH, King RW, and Bock Y (1987) Assessing the effects of atmospheric errors on estimates of relative position obtained from GPS data. *Eos, Transactions American Geophysical Union* 68(16): 286.

Degnan JJ (1993) Millimeter accuracy satellite laser ranging: A review. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Technology (Geodynamics Series)*, vol. 25, pp. 133–162. Washington, DC: American Geophysical Union.

DeMets C, Gordon RG, Argus DF, and Stein S (1990) Current plate motions. *Geophysical Journal International* 101: 425–478.

DeMets C, Gordon RG, Argus DF, and Stein S (1994) Effect of recent revisions of the geomagnetic reversal time scale on estimates of current plate motions. *Geophysical Research Letters* 21: 2191–2194.

Devoti R, Luceri V, Sciarretta C, et al. (2001) The SLR secular gravity variations and their impact on the inference of mantle rheology and lithospheric thickness. *Geophysical Research Letters* 28(5): 855–858.

Dickey JO, Bender PL, Faller JE, et al. (1994) Lunar laser ranging: A continuing legacy of the Apollo Program. *Science* 265: 482–490. http://dx.doi.org/10.1126/science.265.5171.482.

Dixon TH (1991) An introduction to the Global Positioning System and some geological applications. *Reviews of Geophysics* 29(2): 249–276.

Dixon TH, Golombek MP, and Thornton CL (1985) Constraints on Pacific plate kinematics and dynamics with Global Positioning System measurements. *IEEE Transactions on Geoscience and Remote Sensing* GE-23(4): 491–501.

Dixon TH, Norabuena E, and Hotaling L (2003) Paleoseismology and GPS: Earthquake cycle effects and geodetic versus geologic fault slip rates in the eastern California shear zone. *Geology* 31: 55–58.

Dong D and Bock Y (1989) GPS network analysis with phase ambiguity resolution applied to crustal deformation studies in California. *Journal of Geophysical Research* 94(B4): 3949–3966.

Dow JM, Gurtner W, and Schlüter W (2005a) The IGGOS viewed from the Space Geodetic Services. *Journal of Geodynamics* 40(4–5): 375–386.

Dow JM, Neilan RE, and Gendt G (2005b) The International GPS Service (IGS): Celebrating the 10th anniversary and looking to the next decade. *Advances in Space Research* 36(3): 320–326. http://dx.doi.org/10.1016/j.ast.2005.05.125.

Dunn P, Torrence M, Kolenkiewicz R, and Smith D (1999) Earth scale defined by modern satellite ranging observations. *Journal of Geophysical Research* 26(10): 1489–1492.

Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

Dziewonski AM, Hager BH, and O'Connell RJ (1997) Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research* 82: 239–255. http://dx.doi.org/10.1029/JB082i002p00239.

Eubanks TM (1993) Variations in the orientation of the Earth. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*, pp. 1–54. Washington, DC: American Geophysical Union.

Evans AG, Hill RW, Blewitt G, et al. (2002) The Global Positioning System geodesy odyssey, navigation. *Journal of the Institute of Navigation* 49(1): 7–34.

Farrell WE (1972) Deformation of the Earth by surface loads. *Reviews of Geophysics* 10: 761–797.

Fliegel HF and Gallini TE (1996) Solar force modeling of Block IIR Global Positioning System satellites. *Journal of Spacecraft and Rockets* 33(6): 863–866.

Fliegel HF, Gallini TE, and Swift ER (1992) Global Positioning System radiation force model for geodetic applications. *Journal of Geophysical Research* 97: 559–568.

Freymueller JT and Kellogg JN (1990) The extended tracking network and indications of baseline precision and accuracy in the North Andes. *Geophysical Research Letters* 17: 207–210.

Ge M, Gendt G, Dick G, Zhang FP, and Reigber C (2005) Impact of GPS satellite antenna offsets on scale changes in global network solutions. *Geophysical Research Letters* 32(6): L06310. http://dx.doi.org/10.1029/2004GL022224.

Gegout P and Cazenave A (1993) Temporal variations of the Earth gravity field for 1985–1989 derived from LAGEOS. *Geophysical Journal International* 114: 347–359.

Gourmelen N and Amelung F (2005) Post-seismic mantle relaxation in the Central Nevada Seismic Belt. *Science* 310: 1473–1476. http://dx.doi.org/10.1126/science.1119798.

Grafarend EW, Engels J, and Varga P (1997) The spacetime gravitational field of a deforming body. *Journal of Geodesy* 72: 11–30.

Gross RS, Blewitt G, Clarke PJ, and Lavallée (2004) Degree-2 harmonics of the Earth's mass load estimated from GPS and Earth rotation data. *Geophysical Research Letters* 31: L07601. http://dx.doi.org/10.1029/2004GL019589.

Gurtner W, Beutler G, Bauersima I, and Schildknecht T (1985) Evaluation of the GPS carrier difference observations: The BERNESE second generation software package. In: Goad Clyde (ed.), *1st Int. Symp. on Precise Positioning with the Global Positioning System*, Rockville, MD: U.S. Dept. of Commerce.

Hager BH, King RW, and Murray MH (1991) Measurement of crustal deformation using the Global Positioning System. *Annual Reviews of Planetary Science* 19: 351–382.

Haines AJ (1982) Calculating velocity fields across plate boundaries from observed shear rates. *Geophysical Journal of the Royal Astronomical Society* 68: 203–209.

Haines AJ and Holt WE (1993) A procedure for obtaining the complete horizontal motions within zones of distributed deformation from the inversion of strain rate data. *Journal of Geophysical Research* 98: 12057–12082.

Hammond WC (2005) The ghost of an earthquake. *Science* 310: 1440–1442.

Heflin MB, Bertiger WI, Blewitt G, et al. (1992) Global geodesy using GPS without fiducial sites. *Geophysical Research Letters* 19: 131–134.

Henson DJ, Collier EA, and Schneider KR (1985) Geodetic applications of the Texas Instruments TI-4100 GPS navigator. In: Goad C (ed.), *Proceedings of the 1st International Symposium on Precise Positioning with the Global Positioning System*, pp. 191–200. Rockville, MD: U.S. Dept. of Commerce.

Herring TA, Davis JL, and Shapiro II (1990) Geodesy by radio interferometry: The application of Kalman filtering to the analysis of very long baseline interferometry data. *Journal of Geophysical Research* 95: 12561–12581.

Herring TA and Pearlman MR (1993) Future developments and synergism of space geodetic measurement techniques. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Technology*, pp. 21–22. Washington DC: American Geophysical Union.

Hetland EA and Hager BH (2003) Postseismic relaxation across the Central Nevada Seismic Belt. *Journal of Geophysical Research* 108: 2394. http://dx.doi.org/10.1029/2002JB002257.

Hetland EA and Hager BH (2006) The effect of rheological layering on postseismic and interseismic displacements. *Geophysical Journal International* 166: 277–292.

Hill EM and Blewitt G (2006) Testing for fault activity at Yucca Mountain, Nevada, using independent GPS results from the BARGEN network. *Geophysical Research Letters* 33: L14302. http://dx.doi.org/10.1029/2006GL026140.

Holt WE (1995) Flow fields in the Tonga slab determined from the moment tensors of deep earthquakes. *Geophysical Research Letters* 22: 989–992.

Holt WE and Haines AJ (1993) Velocity fields in deforming Asia from the inversion of earthquake-released strains. *Tectonics* 12: 1–20.

Holt WE and Haines AJ (1995) The kinematics of northern South Island New Zealand determined from geologic strain rates. *Journal of Geophysical Research* 100: 17991–18010.

Holt WE, Kreemer C, Haines AJ, et al. (2005) Project helps constrain continental dynamics and seismic hazards. *Eos, Transactions of the American Geophysical Union* 86: 383–387.

Holt WE, Li M, and Haines AJ (1995) Earthquake strain rates and instantaneous relative motions within central and east Asia. *Geophysical Journal International* 122: 569–593.

Holt WE, Ni JF, Wallace TC, and Haines AJ (1991) The active tectonics of the Eastern Himalayan Syntaxis and Surrounding Regions. *Journal of Geophysical Research* 96: 14595–14632.

IERS (2004) McCarthy DD and Petit G (eds.), IERS Conventions 2003, IERS Technical Note 32, Frankfurt am Main: Verlag des Bundesamts für Kartographie und Geodäsie, 127 pp., paperback, ISBN 3-89888-884-3.

Jackson J, Haines AJ, and Holt WE (1992) Determination of the complete horizontal velocity field in the deforming Aegean Sea region from the moment tensors of earthquakes. *Journal of Geophysical Research* 97: 17657–17684.

Jackson JA, Haines AJ, and Holt WE (1994) Combined analysis of strain rate data from satellite laser ranging and seismicity in the Aegean region. *Geophysical Research Letters* 21: 2849–2852.

Jackson JA, Haines AJ, and Holt WE (1995) The accommodation of Arabia–Eurasia plate convergence in Iran. *Journal of Geophysical Research* 100: 15205–15219.

Kaplan GH (1981) The IAU Resolutions on Astronomical Constants, Time Scale and the Fundamental Reference Frame, U. S. Naval Observatory Circular No. 163.

Kedar S, Hajj GA, Wilson BD, and Heflin MB (2003) The effect of the second order GPS ionospheric correction on receiver positions. *Geophysical Research Letters* 30(16): 1829. http://dx.doi.org/10.1029/2003GL017639.

King RW, Abbot RI, Counselman CC, Gourevitch SA, Rosen BJ, and Bock Y (1984) Interferometric determination of GPS satellite orbits. *Eos, Transactions of American Geophysical Union* 65: 853.

King RW and Bock Y (2005) *Documentation for the GAMIT GPS Processing Software Release 10.2*. Cambridge, MA: Mass Inst of Technol.

Kostrov VV (1974) Seismic moment and energy of earthquakes, and seismic flow of rocks. *Izvestiya, Academy of Sciences, USSR, Physics of Solid Earth* 1: 23–44, Eng. Transl.

Kreemer C, Blewitt G, and Hammond WC (2006) Using geodesy to explore correlations between crustal deformation characteristics and geothermal resources. *Geothermal Resources Council Transactions* 30: 441–446.

Kreemer C, Blewitt G, and Maerten F (2006) Co- and postseismic deformation of the 28 March 2005 Nias Mw 8.7 earthquake from continuous GPS data. *Geophysical Research Letters* 33: L07307. http://dx.doi.org/10.1029/2005GL025566.

Kreemer CJ, Haines WE, Holt G Blewitt, and Lavallée D (2000) On the determination of a global strain rate model. *Earth, Planets and Space* 52: 765–770.

Kreemer C, Holt WE, and Haines AJ (2003) An integrated global model of present-day plate motions and plate boundary deformation. *Geophysical Journal International* 154: 8–34.

Kuang D, Rim HJ, Schutz BE, and Abusali PAM (1996) Modeling GPS satellite attitude variation for precise orbit determination. *Journal of Geodesy* 70: 572–580.

Lambeck K (1988) *Geophysical Geodesy: The Slow Deformations of the Earth*. Oxford, UK: Clarendon Press.

Langbein JL, Wyatt F, Johnson H, Hamann D, and Zimmer P (1995) Improved stability of a deeply anchored geodetic monument for deformation monitoring. *Geophysical Research Letters* 22: 3533–3536.

Larson KM and Agnew D (1991) Application of the Global Positioning System to crustal deformation measurement: 1. Precision and accuracy. *Journal of Geophysical Research* 96(B10): 16547–16565.

Larson K, Bodin P, and Gomberg J (2003) Using 1 Hz GPS data to measure deformations caused by the Denali Fault Earthquake. *Science* 300: 1421–1424.

Larson KM, Freymueller JT, and Philipsen S (1997) Global plate velocities from the Global Positioning System. *Journal of Geophysical Research* 102: 9961–9981.

Larson KM, Webb FH, and Agnew D (1991) Application of the Global Positioning System to crustal deformation measurement: 2. The influence of errors in orbit determination networks. *Journal of Geophysical Research* 96(B10): 16567–16584.

Leick A (2004) *GPS Satellite Surveying*, 3rd ed. Hoboken, NJ: Wiley ISBN: 978-0-471-05930-1.

Lichten SM and Border JS (1987) Strategies for high-precision Global Positioning System orbit determination. *Journal of Geophysical Research* 92: 12751–12762.

MacDoran PF (1979) Satellite Emission Radio Interferometric Earth Surveying, SERIES-GPS geodetic system. *Bulletin Geodesique* 53: 117–138.

Mader GL (1999) GPS antenna calibration at the National Geodetic Survey. *GPS Solutions* 3(1): 50–58.

Mader GL and Czopek FM (2002) The Block IIA satellite—Calibrating antenna phase centers. *GPS World* 13(5): 40–46.

Mao A, Harrison CGA, and Dixon TH (1999) Noise in GPS coordinate time series. *Journal of Geophysical Research* 104: 2797–2816.

McCarthy D (ed.) (1996) IERS Conventions, IERS Tech. Note 21. Paris: IERS Central Bureau, Observatory of Paris.

McMillan DS (1995) Atmospheric gradients from very long baseline interferometry observations. *Geophysical Research Letters* 22(9): 1041–1044.

Meehan T, Srinivasan J, Spitzmesser D, et al. (1992) The TurboRogue GPS receiver. In: *Proc. of the 6th Intl, Symp. on Satellite Positioning, Columbus, Ohio, 1*, pp. 209–218.

Minster JB, Altamimi Z, Blewitt G, et al. (2010) *Precise Geodetic Infrastructure: National Requirements for a Shared Resource*. Washington, DC: The National Academies Press, ISBN-10:0-309-15811-7, ISBN-13: 978-0-309-15811-4, 142 pp.

Minster JB and Jordan TH (1978) Present-day plate motions. *Journal of Geophysical Research* 83: 5331–5354.

Mitrovica JX, Davis JL, and Shapiro II (1994) A spectral formalism for computing three-dimensional deformations due to surface loads: 1. Theory. *Journal of Geophysical Research* 99: 7057–7073.

Moore AW (2007) The International GNSS Service: Any questions? *GPS World* 18(1): 58–64.

Nerem RS, Chao BF, Au AY, et al. (1993) Time variations of the Earth's gravitational field from satellite laser ranging to LAGEOS. *Geophysical Research Letters* 20(7): 595–598.

Niell A (1996) Global mapping functions for the atmospheric delay at radio wavelengths. *Journal of Geophysical Research* 101(B2): 3227–3246.

Nikolaidis RM, Bock Y, de Jonge PJ, Agnew DC, and Van Domselaar M (2001) Seismic wave observations with the Global Positioning System. *Journal of Geophysical Research* 106(B10): 21897–21916.

Pearlman MR, Degnan JJ, and Bosworth JM (2002) The international laser ranging service. *Advances in Space Research* 30(2): 135–143.

Peltier WR (1974) The impulse response of a Maxwell Earth. *Reviews of Geophysics and Space Physics* 12: 649–669.
Plag HP, Ambrosius B, Baker TF, et al. (1998) Scientific objectives of current and future WEGENER activities. *Tectonophysics* 294: 177–223.
Plag H-P, Juttner H-U, and Rautenberg V (1996) On the possibility of global and regional inversion of exogenic deformations for mechanical properties of the Earth's interior. *Journal of Geodynamics* 21: 287–309.
Pollitz FF (1997) Gravitational-viscoelastic postseismic relaxation on a layered spherical Earth. *Journal of Geophysical Research* 102: 17921–17941.
Pollitz FF (2003) Post-seismic relaxation theory on a laterally heterogeneous viscoelastic model. *Geophysical Journal International* 154: 1–22.
Prescott WH, Davis JL, and Svarc JL (1989) Global Positioning System measurements for crustal deformation: Precision and accuracy. *Science* 244: 1337–1340.
Ray J (ed.) (1998) IERS Analysis Campaign to Investigate Motions of the Geocenter, International Earth Rotation Service Tech. Note 25, Obs. Paris.
Rogers AEE, Knight CA, Hinteregger HF, et al. (1978) Geodesy by radio interferometry: Determination of a 1.24-km base line vector with ~5-mm repeatability. *Journal of Geophysical Research* 83: 325–334.
Rothacher M, Beutler G, Gurtner W, Schildknecht T, and Wild U (1990) *BERNESE GPS Software Version 3.2*. Switzerland: Printing Office Univ of Berne.
Rothacher M, Springer TA, Schaer S, and Beutler G (1998) Processing Strategies for Regional Networks. In: Brunner FK (ed.) *Advances in Positioning and Reference Frames, IAG Symposium No. 118,* pp. 93–100.
Rummel R, Rothacher M, and Beutler G (2005) Integrated Global Geodetic Observing System (IGGOS)—Science rationale. *Journal of Geodynamics* 40(4–5): 357–362.
Sabadini R, Di Donato G, Vermeersen LLA, Devoti R, Luceri V, and Bianco G (2002) Ice mass loss in Antarctica and stiff lower mantle viscosity inferred from the long wavelength time dependent gravity field. *Geophysical Research Letters* 29(10).
Savage JC (1983) Strain accumulation in the Western United States. *Annual Review of Earth and Planetary Sciences* 11: 11–41. http://dx.doi.org/10.1145/annurev.ea.11.0501083.000303.
Schlüter W, Himwich E, Nothnagel A, Vandenberg N, and Whitney A (2002) IVS and its important role in the maintenance of the global reference systems. *Advances in Space Research* 30(2): 145–150.
Schmid R and Rothacher M (2003) Estimation of elevation-dependent satellite antenna phase center variations of GPS satellites. *Journal of Geodesy* 77: 440–446. http://dx.doi.org/10.1007/s00190-003-0339-0.
Schmid R, Rothacher M, Thaller D, and Steigenberger P (2005) Absolute phase center corrections of satellite and receiver antennas impact on GPS solutions and estimation of azimuthal phase center variations of the satellite antenna. *GPS Solutions* 9(4): 283–293.
Schupler BR, Allshouse RL, and Clark TA (1994) Signal characteristics of GPS user antennas. *Journal of the Institute of Navigation* 41: 277–295.
Segall P and Davis JL (1997) GPS applications for geodynamics and earthquake studies. *Annual Reviews of Earth and Planetary Sciences* 25: 301–336.
Sella G, Dixon T, and Mao A (2002) REVEL: A model for recent plate velocities from space geodesy. *Journal of Geophysical Research* 107(B4). http://dx.doi.org/10.1029/2000JB000033.
Shapiro II (1964) Fourth test of general relativity. *Physical Review Letters* 13(26): 789–791. http://dx.doi.org/10.1103/PhysRevLett.13.789.
Shen-Tu B, Holt WE, and Haines AJ (1995) Intraplate deformation in the Japanese Islands: A Kinematic study of intraplate deformation at a convergent plate margin. *Journal of Geophysical Research* 100: 24275–24293.
Shen-Tu B, Holt WE, and Haines AJ (1999) The kinematics of the western United States estimated from Quaternary rates of slip and space geodetic data. *Journal of Geophysical Research* 104: 28927–28955.
Silver PG, Bock Y, Agnew D, et al. (1999) A plate boundary observatory. *IRIS Newsletter* 16: 3–9.
Smith DE and Turcotte DL (eds.) (1993) *Contributions of Space Geodesy to Geodynamics: Technology*. Washington, DC: American Geophysical Union.
Smith KD, von Seggern D, Blewitt G, et al. (2004) Evidence for deep magma injection beneath Lake Tahoe, Nevada-California. *Science* 305: 1277–1280. http://dx.doi.org/10.1126/science.1101304.
Sovers OJ and Border JS (1987) *Observation Model and Parameter Partials for the JPL Geodetic Software, "GPSOMC"*. Pasadena, CA: JPL Pub Jet Propulsion Laboratory, 87–21.
Steigenberger P, Rothacher M, Dietrich R, Fritsche M, Rülke A, and Vey S (2006) Reprocessing of a global GPS network. *Journal of Geophysical Research* 111: B05402. http://dx.doi.org/10.1029/2005JB003747.
Stein S (1993) Space geodesy and plate motions. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Crustal Dynamics*, pp. 5–20. Washington, DC: American Geophysical Union.
Su W, Woodward RL, and Dziewonski AD (1994) Degree 12 model of shear velocity heterogeneity in the mantle. *Journal of Geophysical Research* 99: 6945–6980.
Swift ER (1985) NSWC's GPS orbit/clock determination system. In: *Proc. of the 1st Symp. on Prec. Positioning with the Global Positioning System, U.S. Dept. of Commerce, Rockville, MD* .
Tapley BD, Schutz BE, Eanes RJ, Ries JC, and Watkins MM (1993) LAGEOS laser ranging contributions to geodynamics, geodesy and orbital dynamics. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*, pp. 147–174. Washington, DC: American Geophysical Union.
Tavernier G, Fagard H, Feissel-Vernier M, et al. (2005) The International DORIS Service, IDS. *Advances in Space Research* 36(3): 333–341.
Thomas JB (1988) *Functional Description of Signal Processing in the Rogue GPS Receiver*. Pasadena, CA: JPL, pp 88–15.
Tinnon M, Holt WE, and Haines AJ (1995) Velocity gradients in the northern Indian Ocean inferred from earthquake moment tensors and relative plate velocities. *Journal of Geophysical Research* 100: 24315–24329.
Torge W (2001) *Geodesy*, 3rd ed. Berlin, Germany: Gruyte ISBN: 978-3110170726, 400 pp.
Tralli DM, Dixon TH, and Stephens SA (1988) The effect of wet tropospheric path delays on estimation of geodetic baselines in the Gulf of California using the Global Positioning System. *Journal of Geophysical Research* 93(B6): 6545–6557.
Tralli DM and Lichten SM (1990) Stochastic estimation of tropospheric path delays in Global Positioning System geodetic measurements. *Bulletin Geodesique* 64: 127–159.
Tregoning P and Herring TA (2006) Impact of a priori zenith hydrostatic delay errors on GPS estimates of station heights and zenith total delays. *Geophysical Research Letters* 33: L23303. http://dx.doi.org/10.1029/2006GL027706.
Truehaft RN and Lanyi GE (1987) The effects of the dynamic wet troposphere on radio interferometric measurements. *Radio Science* 22: 251–265.
Van Dam T, Blewitt G, and Heflin MB (1994) Atmospheric pressure loading effects on GPS coordinate determinations. *Journal of Geophysical Research* 99: 23939–23950.
Van Dam TM, Wahr J, Milly PCD, Shmakin AB, Blewitt G, and Larson KM (2001) Crustal displacements due to continental water loading. *Geophysical Research Letters* 28: 651–654.
Wahr JM (1981) Body tides on an elliptical, rotating, elastic and oceanless Earth. *Geophysical Journal* 64: 677–703.
Watkins MM and Eanes RJ (1997) Observations of tidally coherent diurnal and semidiurnal variations in the geocenter. *Geophysical Research Letters* 24: 2231–2234.
Wdowinski S, Bock Y, Zhang J, Fang P, and Genrich J (1997) Southern California Permanent GPS Geodetic Array; spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake. *Journal of Geophysical Research* 102(B8): 18057–18070.
Webb FH and Zumberge JF (1993) *An Introduction to GIPSY-OASIS II*. Pasadena, CA: Jet Propulsion Laboratory, JPL Publication D-11088.
Wernicke BP, Friedrich AM, Niemi NA, Bennett RA, and Davis JL (2000) Dynamics of plate boundary fault systems from Basin and Range Geodetic Network (BARGEN) and geologic data. *GSA Today* 10(11): 1–7.
Williams SDP (2003) The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. *Journal of Geodesy* 76: 483–494. http://dx.doi.org/10.1007/s00190-002-0283-4.
Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106(E11): 27933–27968.
Williams JG, Tunyshev S, and Boggs DH (2004) Progress in lunar laser ranging tests of relativistic gravity. *Physical Review Letters* 93. http://dx.doi.org/10.1103/PhysRevLett.93.261101.
Wilson P and Reinhart E (1993) The Wegener-Medlas Project preliminary results on the determination of the geokinematics of the Eastern Mediterranean. In: Smith DE and Turcotte DL (eds.) *Contributions of Space Geodesy to Geodynamics: Crustal Dynamics*, pp. 299–310. Washington, DC: American Geophysical Union.
Wu JT, Wu SC, Hajj GA, Bertiger WI, and Lichten SM (1993) Effects of antenna orientation on GPS carrier phase. *Manuscripta Geodaetica* 18: 91–93.
Wyatt F (1982) Displacement of surface monuments; horizontal motion. *Journal of Geophysical Research* 87: 979–989.
Ziebart M, Cross P, and Adhya S (2002) Modeling photon pressure, the key to high-precision GPS satellite orbits. *GPS World* 13(1): 43–50.
Zumberge JF, Heflin MB, Jefferson DC, Watkins MM, and Webb FH (1997) Precise point positioning for the efficient and robust analysis of GPS data from large networks. *Journal of Geophysical Research* 102(B3): 5005–5018.

3.12 Interferometric Synthetic Aperture Radar Geodesy

M Simons and PA Rosen, California Institute of Technology, Pasadena, CA, USA

© 2015 Elsevier B.V. All rights reserved.

3.12.1	**Introduction**	339
3.12.1.1	Motivation	339
3.12.1.2	History and Overview	340
3.12.2	**InSAR**	344
3.12.2.1	The Interferogram	346
3.12.2.2	Interferometric Baseline and Height Reconstruction	346
3.12.2.3	Differential Interferometry	348
3.12.2.4	Phase Unwrapping	349
3.12.2.5	Correlation	349
3.12.2.6	Processing Data	351
3.12.3	**InSAR-Related Techniques**	351
3.12.3.1	Wide-Swath Interferometry	351
3.12.3.2	Permanent Scatterers and Time-Series Analysis	354
3.12.3.3	Speckle Tracking and Pixel Tracking	355
3.12.3.4	Other Advances	356
3.12.4	**A Best-Practices Guide to Using InSAR and Related Observations**	357
3.12.4.1	Interferometric Baseline Errors	358
3.12.4.2	Propagation Delays	358
3.12.4.3	A Simple Data Covariance Model for Multiple Interferograms	361
3.12.4.4	Impact of Solid Earth and Ocean Tides	361
3.12.4.5	Stacking Single Component Images	361
3.12.4.6	InSAR Time Series	362
3.12.4.7	Vector Displacements	364
3.12.4.8	The Luxury of Sampling: Rationale and Approach	364
3.12.4.9	Decorrelation as Signal	366
3.12.5	**The Link Between Science and Mission Design**	367
Appendix A		370
Acknowledgments		382
References		382

3.12.1 Introduction

3.12.1.1 Motivation

In 1993, Goldstein et al. (1993) presented the first satellite-based interferometric synthetic aperture radar (InSAR) map showing large strains of the Earth's solid surface – in this case, the deforming surface was an ice stream in Antarctica. The same year, Massonnet et al. (1993) showed exquisitely detailed and spatially continuous maps of surface deformation associated with the 1992 M_w 7.3 Landers earthquake in the Mojave Desert in southern California. These papers heralded a new era in geodetic science, whereby we can potentially measure three-dimensional surface displacements with nearly complete spatial continuity, from a plethora of natural and human-induced phenomena. In addition to the increase spatial resolution afforded by InSAR, the increasing depth of image archives now permit the construction of detailed time series of ground deformation (e.g., Figure 1). An incomplete list of targets to date includes all forms of deformation on or around faults (interseismic, aseismic, coseismic, and postseismic) aimed at constraining the rheological properties of the fault and surrounding crust, detection and quantification of changes in active magma chambers aimed at understanding a volcano's plumbing system, slow-moving landslides, the mechanics of glaciers and temporal changes in glacier flow with obvious impacts on assessments of climate change, and the impact of seasonal and anthropogenic changes in aquifers (e.g., Figure 1). Beyond detection of coherent surface deformation, InSAR can also provide unique views of surface disruption, through measurements of interferometric decorrelation, that could potentially aid the ability of emergency responders to respond efficiently to many natural disasters.

Today, spaceborne InSAR enjoys widespread application, in large part because of the availability of suitable globally acquired synthetic aperture radar (SAR) data from the ERS-1, ERS-2, and Envisat satellites operated by the European Space Agency, JERS-1 and ALOS satellites operated by the National Space Development Agency of Japan, Radarsat-1 and Radarsat-2 operated by the Canadian Space Agency, TerraSAR-X and TanDEM-X operated by the German Space Agency, and the COSMO-SkyMed constellation operated by the Italian Space Agency, with new and follow-on systems in the works. As more and more radar data become available from international civilian radar satellites and as scientific demands become greater on the use of these data, including extraction of ever more subtle and well-calibrated geophysical signals, it is

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00061-0

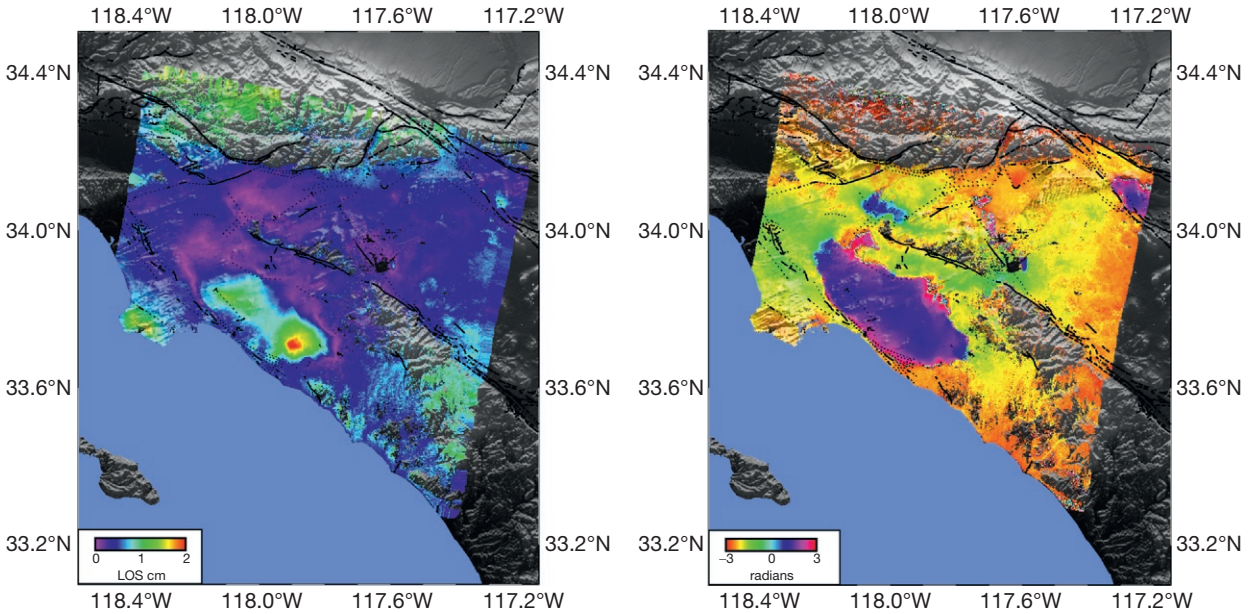


Figure 1 An example of using large InSAR data sets to image time-dependent surface deformation of the Los Angeles Basin (California). This result uses 105 ERS 1/2 images plus 60 Envisat images to make a total of 881 interferograms spanning different time periods between June 1992 and September 2010. The ensemble of interferograms is used in a time-series analysis that accounts for slowly changing temporal deformation plus seasonal terms (Agram et al., 2013). Shown here are the amplitude (left) and phase (right) of the seasonal terms in color superimposed on a shaded relief rendition of topography. Black lines indicate major faults. Sharp boundaries in the phase image outline the major aquifer systems. Similar results have been found previously (e.g., Argus et al., 2005; Bawden et al., 2001; Lanari et al., 2004b; Watson, 2002). Figure courtesy of Piyush Agram.

essential to understand the characteristics of the image, how they are processed, and how that processing can affect the interpretation of the image data.

In this chapter, we aim to provide a review of the basic theory of InSAR for geodetic applications. Numerous review articles and books on the topic of InSAR already exist (e.g., Bürgmann et al., 2000; Hanssen, 2001; Massonnet and Feigl, 1995; Rosen et al., 2000), and our goal is not to repeat these works more than necessary. Instead, we attempt to provide an overview of what data, processing, and analysis schemes are currently used and a glimpse of what the future may hold. As part of this discussion, we present our experience-based view of what constitutes best practices for use of InSAR observations in geodetic modeling. Finally, we provide a guide to the interplay between different mission design parameters and their relationship to the character of the resulting observations, which may aid the reader in how to approach the disparate array of available data for their problem of interest.

The last few years have seen an explosion in the literature describing the development and the use of InSAR methods and observations. Given the limited amount of space available, we mention what we feel to be a representative subset of these studies. This review borrows heavily from our previous work with many colleagues, and where appropriate, we point the reader to the original sources for a more complete discussion. Much of the SAR processing discussion is derived and simplified from Rosen et al. (2000), although here, this discussion is augmented to include a variety of more recent techniques including persistent scatterers, ScanSAR interferometry, and pixel tracking.

3.12.1.2 History and Overview

Operating at microwave frequencies, SAR systems provide unique images representing the electrical and geometric properties of a surface in nearly all weather conditions. Since they provide their own illumination, SARs can image in daylight or at night. SAR mapping systems typically operate on airborne or spaceborne platforms following a linear flight path, as illustrated in Figure 2. Raw image data are collected by transmitting a series of coded pulses from an antenna illuminating a swath offset from the flight track. The echo of each pulse is recorded during a period of reception between the transmission events. When a number of pulses are collected, it is possible to perform two-dimensional matched-filter compression on a collection of pulse echoes to focus the image. This process is known as SAR because in the along-track, or azimuth, direction, a large virtual aperture is formed by coherently combining the collection of radar pulses received as the SAR moves along in its flight path (Raney, 1971). Although the typical physical length of a SAR antenna is on the order of meters, the synthesized aperture length can be of the order of kilometers. Because the image is acquired from a side-looking vantage point (to avoid left-side/right-side ambiguities), the radar image is geometrically distorted relative to the ground coordinates (Figure 2).

Figure 3 illustrates the InSAR system concept. By coherently combining the signals from two antennas, the interferometric phase difference between the received signals can be formed for each imaged point. In this scenario, the phase difference is essentially related to the geometric path length difference to the image point. In the left panel of Figure 3, the path

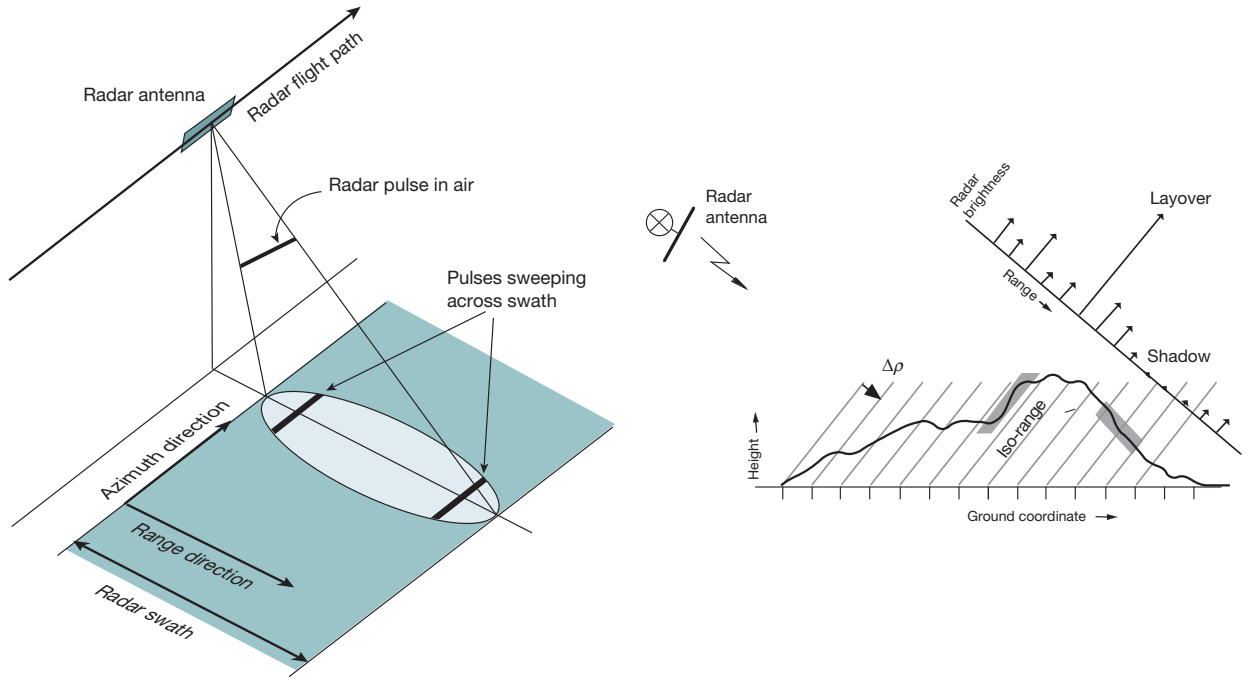


Figure 2 (Left) Typical imaging scenario for a SAR system. The platform carrying the SAR instrument follows a curvilinear track known as the 'along-track' or 'azimuth' direction. The radar antenna points to the side, imaging the terrain below. The distance from the aperture to a target on the surface in the look direction is known as the 'cross track' or 'range direction' and is terrain-dependent. The radar sends a pulse that sweeps through the antenna beam, effectively returning the integrated backscatter over the pulse and azimuth beam extent at any given instant. The azimuth extent can be many kilometers. Matched filtering creates fine resolution in range. Synthetic aperture processing creates fine resolution in azimuth. (Right) The three-dimensional world is collapsed to two dimensions in conventional SAR imaging. After image formation, the radar return is resolved into an image in range-azimuth coordinates. This figure shows a profile of the terrain at constant azimuth, with the radar flight track into the page. The profile is cut by curves of constant range, spaced by the range resolution of radar, defined as $\Delta\rho = c/2\Delta f_{BW}$, where c is the speed of light and Δf_{BW} is the range bandwidth of the radar. The backscattered energy from all surface scatterers within a range resolution element contributes to the radar return for that element.

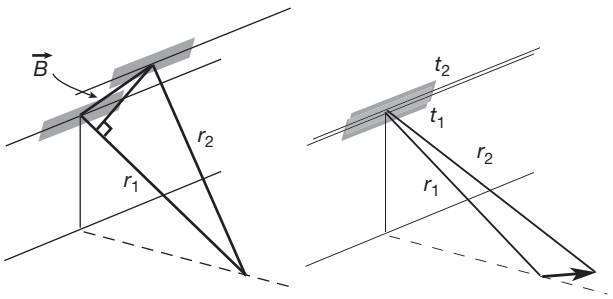


Figure 3 Interferometric SAR for topographic mapping uses two apertures separated by a 'baseline,' B, to image the surface. The phase difference between the apertures for each image point, along with the range and knowledge of the baseline, can be used to infer the precise shape of the imaging triangle to derive the topographic height of the image point. At left, a range difference exists because the scene is viewed from two different vantage points. This is described by a shift in the point target response as presented in the text. At right, a range difference is generated by a change in the position of the scene from one time to the next, imaged from the same vantage point. This range difference is described by a scene shift, not a point target response shift; the mathematics is the same but for a sign change.

difference depends on the topography. With the knowledge of the interferometer geometry, the phase difference can be converted into an altitude for each image point. In essence, the phase difference provides a third measurement, in addition to the along and cross track location of the image point, or 'target,' to allow a reconstruction of the three-dimensional location of the targets.

The InSAR approach for topographic mapping is similar to the conventional stereoscopic approach. In stereoscopy, a pair of images of the terrain are obtained from two displaced imaging positions. The 'parallax' obtained from the displacement allows the retrieval of topography because targets at different heights are displaced relative to each other in the two images by an amount related to their altitudes (Rosen et al., 2000). InSAR differs from stereoscopy in that the 'parallax' measurements between the SAR images are obtained by measuring the phase difference between the signals received by two InSAR antennas. These phase differences can be used to determine the angle of the target relative to the baseline of the interferometric SAR directly. The accuracy of the InSAR parallax measurement is typically several millimeters to centimeter, being a fraction of the SAR wavelength, whereas the parallax measurement accuracy of the stereoscopic approach is usually on the order of the resolution of the imagery (several meters or more).

Typically, the postspacing of the InSAR topographic data is comparable to the fine spatial resolution of SAR imagery, while the altitude measurement accuracy generally exceeds stereoscopic accuracy at comparable resolutions. The registration of the two SAR images for the interferometric measurement, the

retrieval of the interferometric phase difference, and the subsequent conversion of the results into digital elevation models (DEMs) of the terrain can be highly automated, an advantageous aspect of the InSAR approach. As discussed in the succeeding sections, the performance of InSAR systems is largely understood both theoretically and experimentally. These developments have led to airborne and spaceborne InSAR systems for routine topographic mapping. The InSAR technique just described, using two apertures on a single platform, is often called 'cross track interferometry' in the literature. Other terms are 'single-track' and 'single-pass' interferometry.

Goldstein and Zebker (1987) advanced another interferometric SAR technique to measure surface motion by imaging the surface at multiple times. The time separation between the imaging can be a fraction of a second to years. The multiple images can be thought of as 'time-lapse' imagery. Movement of a target will be detected by comparing the images. Unlike conventional schemes in which motion is detected only when the targets move more than a significant fraction of the resolution of the imagery, this technique measures the phase differences of the pixels in each pair of the multiple SAR images. If the flight path and imaging geometries of all the SAR observations are identical, any interferometric phase difference is due to changes over time of the SAR system clock, variable propagation delay, or surface motion in the direction of the radar line of sight.

In the first application of this technique described in the open literature, Goldstein and Zebker (1987) augmented a conventional airborne SAR system with an additional aperture, separated along the length of the aircraft fuselage from the conventional SAR antenna. Given an antenna separation of roughly 20 m and an aircraft speed of about 200 m s^{-1}, the time between target observations made by the two antennas was about 100 ms. Over this time interval, clock drift and propagation delay variations are negligible. This system measured tidal motions in the San Francisco Bay Area with an accuracy of several centimeters per second (Goldstein and Zebker, 1987). This technique has been dubbed 'along-track interferometry' (ATI) because of the arrangement of two antennas along the flight track on a single platform. In the ideal case, there is no cross track separation of the apertures and therefore no sensitivity to topography.

ATI is merely a special case of 'repeat-track interferometry' (RTI), which can be used to generate topography and motion. The orbits of several spaceborne SAR satellites have been controlled in such a way that they nearly retrace themselves after several days. Aircraft can also be controlled to accurately repeat flight paths. If the repeat flight paths result in a cross track separation and the surface has not changed between observations, then the repeat-track observation pair can act as an interferometer for topography measurement. For spaceborne systems, RTI is usually termed 'repeat-pass interferometry' in the literature.

If the flight track is repeated perfectly such that there is no cross track separation, then there is no sensitivity to topography, and radial motions can be measured directly as with an ATI system. However, since the temporal separation between the observations is typically days to many months or years, the ability to detect small radial velocities is substantially better than the ATI system described in the preceding text. The first demonstration of RTI for velocity mapping was a study of the Rutford Ice Stream in Antarctica (Goldstein et al., 1993). The radar aboard the ERS-1 satellite obtained several SAR images of the ice stream with near-perfect retracing so that there was no topographic signature in the interferometric phase permitting measurements of the ice stream flow velocity of the order of 1 m $year^{-1}$ (or 3×10^{-8} m s^{-1}) observed over a few days (Goldstein et al., 1993).

Most commonly for repeat-track observations, the track of the sensor does not repeat itself exactly, so the interferometric time-separated measurements generally comprise the signature of topography and of radial motion or surface displacement. The approach for reducing these data into velocity or surface displacement by removing topography is generally referred to as 'differential interferometric SAR.'

Goldstein et al. (1988) conducted the first proof-of-concept experiment for spaceborne InSAR using imagery obtained by the SeaSAT mission. In the latter portion of that mission, the spacecraft was placed into a near-repeat orbit every 3 days. Gabriel et al. (1989) used data obtained in an agricultural region in California, United States, to detect what was interpreted to be surface elevation changes in some of the agricultural fields of the order of several cm over approximately a 1-month period (**Figure 4**). By comparing the areas with the detected surface elevation changes with irrigation records, they concluded that these areas were irrigated in between the observations, causing small elevation changes from increased soil moisture. Gabriel et al. (1989) were actually looking for the

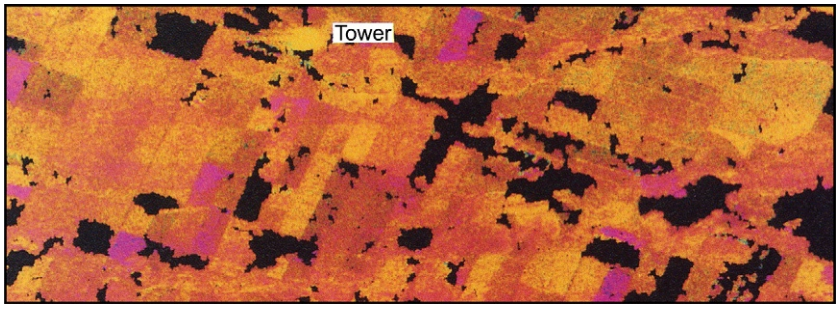


Figure 4 A portion of the original deformation map from the first proof-of-concept experiment for spaceborne InSAR using imagery obtained by the SeaSAT mission (Gabriel et al., 1989). Colored patches delineated by field boundaries indicate possible swelling/compaction of the fields due to moisture changes. The image is approximately 15 km across.

deformation signature of a small earthquake, but the surface motion was too small to detect. These early studies were then followed by the aforementioned seminal applications to glacier flow and earthquake-induced surface deformation (Goldstein et al., 1993; Massonnet et al., 1993).

All civilian InSAR capable satellites to date have been right-looking in near-polar sun-synchronous orbits with the ability to observe a particular location on the Earth on both ascending and descending orbit passes. With a single satellite, it is therefore possible to obtain geodetic measurements corresponding to at least two components of the underlying three-component surface deformation field. The variety of available viewing geometries can be increased if a satellite has both left and right-looking capability. Similarly, neighboring orbital tracks with overlapping beams at different incidence angles can also provide diversity of viewing geometry.

In an ideal mission scenario (see Section 3.12.5), observations from a given viewing geometry will be acquired frequently and for a long period of time to provide a dense archive for InSAR analysis. The frequency of imaging is key in order to provide optimal time resolution of a given phenomena and to provide the ability to combine multiple images to detect small signals. Of course, many processes of interest are not predictable in time, thus we must continuously image the Earth in a systematic fashion in order to provide recent 'before' images. For a given target, not all acquisitions are necessarily viable for InSAR purposes. The greatest nemesis for InSAR geodesy comes from incoherent phase returns between two image acquisitions. This incoherence can be driven by observing geometry (i.e., the baseline is too large) or by physical changes of the Earth's surface (e.g., snowfall). Thus, any InSAR study begins with an assessment of the available image archive. Figure 5 uses an example from the ERS-1 and ERS-2 image archive to illustrate how one would go about choosing images for InSAR processing assuming you wanted to make all available pairs that were not decorrelated due to large baselines, snow, or temporal separation that was too large. Several operating missions have sufficiently tight control on the satellite orbit that baseline selection is not as much an issue as in the past, and it is expected that future missions will only improve on this trend.

That InSAR can take advantage of the satellite's perspective of the world permits one to view large areas of the Earth's surface quickly and efficiently. In solid Earth geophysics, we are frequently interested in rare and extreme events (e.g., earthquakes, volcanic eruptions, and glacier surges). Therefore, if we want to capture these events and their natural variability, we cannot simply rely on dense instrumentation of a few select areas; instead, we must embrace approaches that allow global access. Given easy access to data (which is not always the case), this inherently global perspective provided by satellite-based InSAR also allows one the luxury of going on geodetic fishing trips, whereby one essentially asks 'I wonder if. . .?,' in search of the unexpected (e.g., Figure 6). In essence, we must not limit ourselves to hypothesis testing, but rather, we must also tap the inherently exploratory power of InSAR.

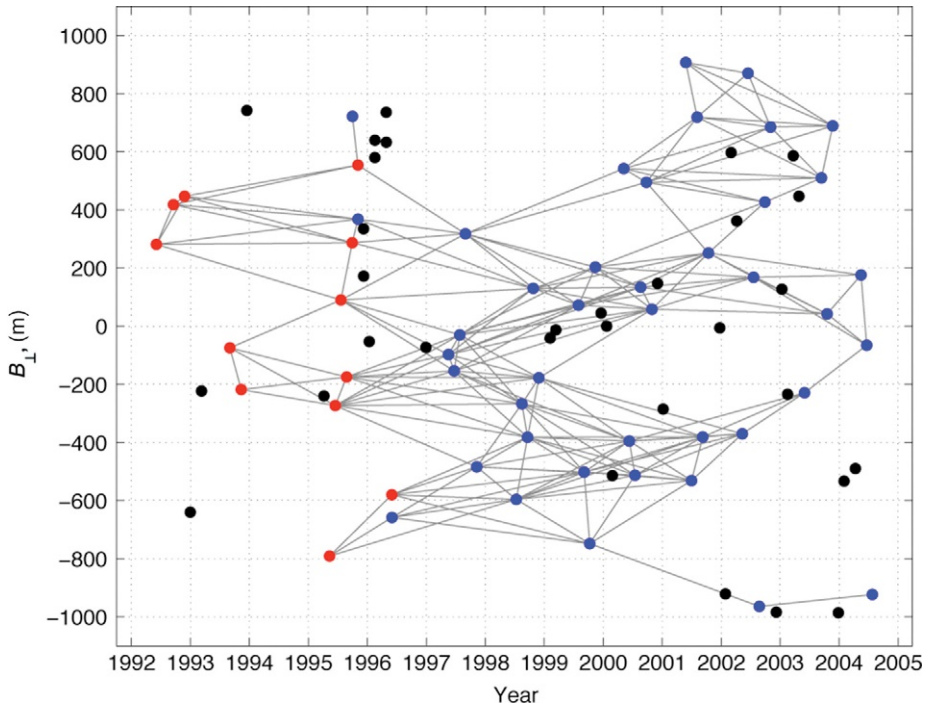


Figure 5 A typical perpendicular component of baseline, $B_\perp$, versus time plot used to select data for InSAR purposes. Here, each dot represents a different acquisition of the same scene by the ERS-1 or ERS-2 satellites. This example is for track 485 and frame 2853 corresponding to the Long Valley Caldera region of California. Lines connecting the dots represent interferograms that are likely to work. We have chosen to indicate only pairs that satisfy the constraints: $B_\perp < 250$ m, neither scene occurs in winter when snowfall and rain cause decorrelation, and $\Delta T < 3{:}5$ years. For viable scenes, red and blue dots indicate ERS-1 and ERS-2, respectively.

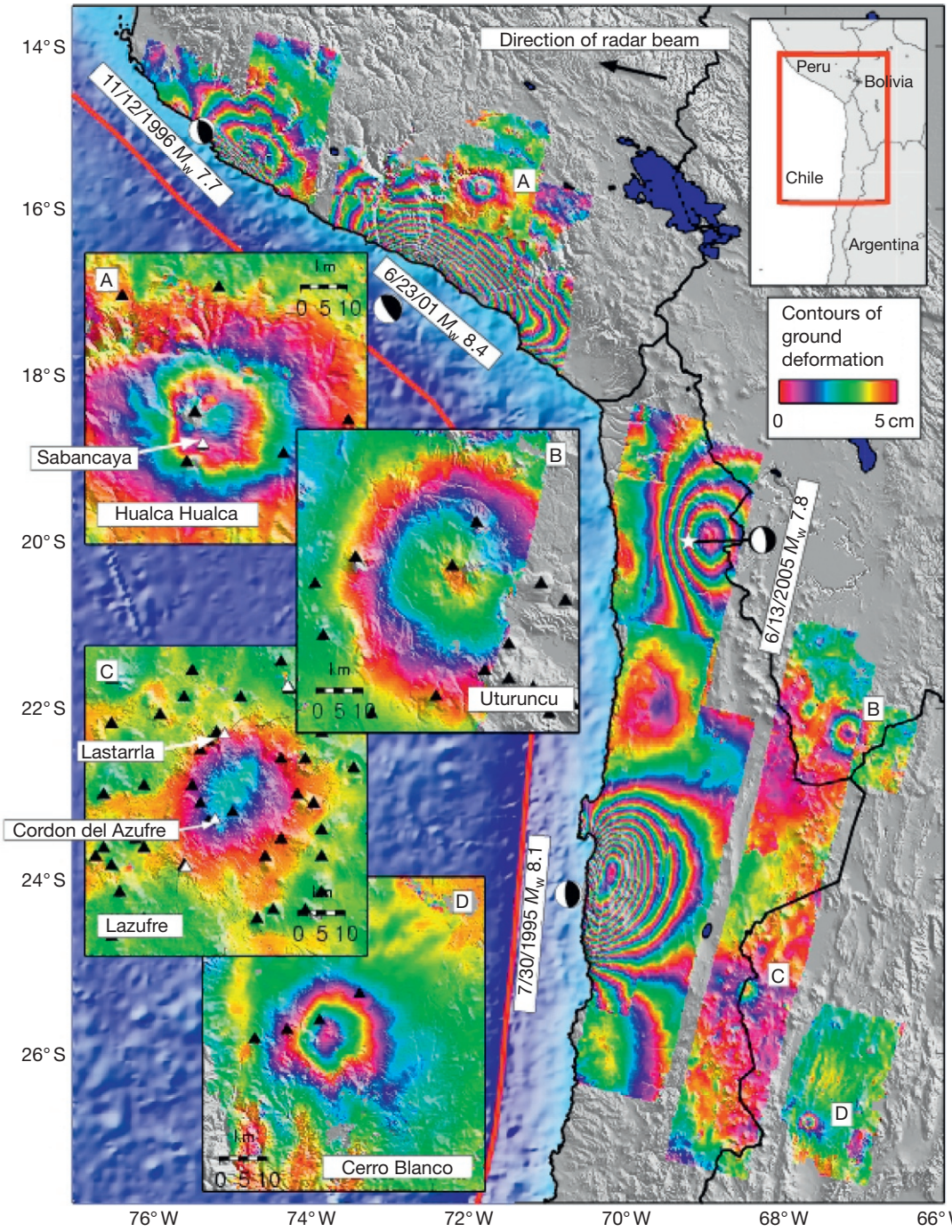


Figure 6 Mosaic of interferograms showing surface deformation associated with subsurface magma migration (inset panels A–D and associated tags in main image) and from three megathrust earthquakes and one intraslab earthquake all with M_w between 7.7 and 8.5. These interferograms are all from right-looking descending orbit C-band radars (ERS-1, ERS-2, and Envisat). The phase has been unwrapped from the natural fringe rate and rewrapped at 5 cm per fringe. The subduction trench is indicated by the red barbed line and the centroid moment tensors with the beach balls. Note the original survey for volcano deformation is a clear example of the 'fishing trip' approach enabled by InSAR. Figure courtesy of Matt Pritchard.

3.12.2 InSAR

Interferometry relies on the constructive and destructive interference of electromagnetic waves from sources at two or more vantage points to infer something about the sources or the relative path length of the interferometer. For InSAR, the interference pattern is constructed from two complex-valued SAR images, and interferometry is the study of the phase difference between two images – acquired from different vantage points, different times, or both.

Appendix A.1 describes the SAR technique, developing a model for the image one would obtain from an idealized

surface that consists entirely of a single reflective point, known as a point target, and then further considering the image effects of natural surfaces. Fine image resolution is achieved in the cross track, or range, direction by transmitting a coded waveform with sufficient bandwidth. Matched-filter compression of each received signal pulse then recovers the range resolution. In the along-track, or azimuth, direction, a SAR forms a large synthetic aperture by coherently combining an ensemble of the radar pulses received as the SAR moves along in its flight path. Matched filtering then focuses the image in azimuth.

In the appendix, it is shown that there is a phase term exp $\{-j2kR\}$, where $k=2\pi/\lambda$ is the wavenumber and λ is the radar wavelength, that characterizes the two-way propagation distance from the radar sensor to the point target and back again. For a general surface, there is an additional phase term contributed by each surface scatterer. The net phase of each image point is the sum of these two terms: the intrinsic phase of the surface, which tends to be random, and the propagation phase term.

A resolution element can be represented as a complex phasor of the coherent backscatter from the scattering elements on the ground and the propagation phase delay, as illustrated in Figure 7. The backscatter phase delay is the net phase of the coherent sum of the contributions from all elemental scatterers in the resolution element, each with their individual backscatter phases and their differential path delays relative to a reference surface normal to the radar look direction.

Radar images observed from two nearby antenna locations have resolution elements with nearly the same complex phasor return, but with a different propagation phase delay. In interferometry, the complex phasor information of one image is multiplied by the complex conjugate phasor information of the second image to form an interferogram, effectively canceling the common backscatter phase in each resolution element, but leaving a phase term proportional to the differential path delay. Ignoring the slight difference in backscatter phase of the surface observed from two different vantage points treats each resolution element as a point scatterer.

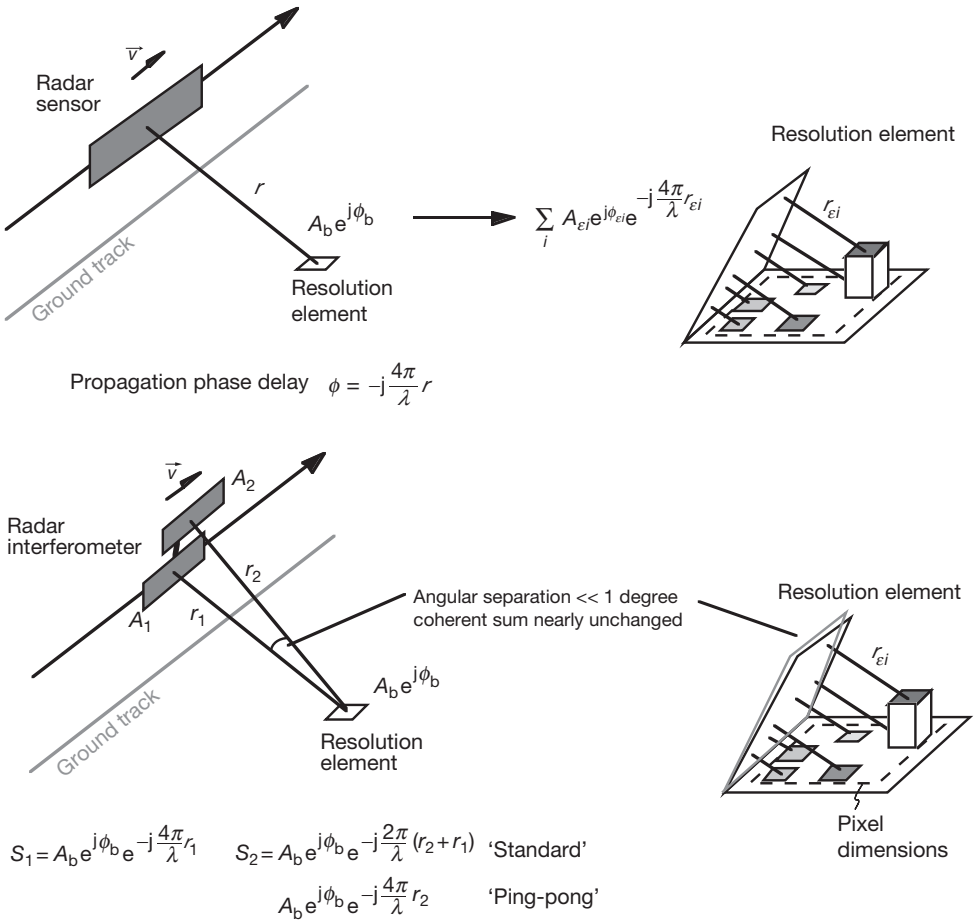


Figure 7 Illustration of the elements of phase in SAR and InSAR observations. (Top) Each resolution element in a radar image has a complex value. Its amplitude, A_b, is related to the electrical reflectivity of the surface and its roughness. Its phase is the sum of the propagation phase delay, $-j(4\pi/\lambda)r$, and the surface backscatter phase, ϕ_b. The surface contribution, $A_b e^{j\phi_b}$, can be modeled as the coherent sum of contributions from elemental complex scatterers, $A_{\varepsilon i} e^{j\phi\varepsilon i}$, each with their individual path delays, $r_{\varepsilon i}$, relative to a common reference. Since the arrangement of scatterers within a resolution element is generally random, the amplitude A_b and phase ϕ_b are also random from element to element. (Bottom) For InSAR, when two observations are made from nearly the same viewing geometry, the backscatter values, $A_{\varepsilon i} e^{j\phi\varepsilon i}$, are nearly the same in each observation, and the phase difference, $\phi_1 - \phi_2$, is essentially the difference in the path delay $-j(4\pi/\lambda)(r_1 - r_2)$.

3.12.2.1 The Interferogram

Given this conceptual framework, and with reference to the variables defined in the preceding text, the 'interferogram' can be defined as the following complex quantity:

$$I(x', r') = |\Gamma(x', r')|^2 e^{j4\pi\delta r(x', r')/\lambda} \quad [1]$$

where $\Gamma(x', r') = A_b(x', r')e^{j\phi_b(x', r')}$ is the complex reflectivity of the surface described in Figure 7 and $\delta r(x', r')$ is the difference of the range from a given patch of ground to the sensor at the two observation vantage points. As noted in the preceding text, each image has phase components ϕ_1 and ϕ_2 composed of ϕ_b and their respective propagation distance $-j4\pi r_1/\lambda$ and $-j4\pi r_1/\lambda$, such that the interferogram's phase, or the 'interferometric phase,' is given by

$$\begin{aligned}\phi_I &= \phi_1 - \phi_2 \\ &= \frac{4\pi}{\lambda}(r_2 - r_1) \\ &= \frac{4\pi}{\lambda}\delta r\end{aligned} \quad [2]$$

A more complete derivation of eqn [1] can be found in the appendix, illustrating some of the subtleties with regard to radar imaging.

The range δr is a relative measure, so the range difference can be induced if the ground stays fixed and the observation points are in different locations or vice versa. Consider the case where the range difference δr arises from a cross track separation of two observation points, as illustrated in Figure 3 at the left. There is a clear dependence on the relative lengths of the two sides of the triangle on the height of the surface, which in general is not known a priori. Thus, δr is not known exactly to align the reflectivities to form the interferogram. However, in practical systems, one can match the reflectivity estimates in the two SAR observations to within sufficient accuracy (generally much better than the image resolution) to derive a sufficient estimate of δr for alignment. Once formed, the interferogram for the cross track interferometer then contains a record of the variability of the height of the surface. It is possible to invert the phase to reconstruct the height. It turns out it can be done quickly and efficiently and is a powerful tool for topographic mapping. Note that the sign of the propagation phase delay is set by the desire for consistency between the Doppler frequency, f_D, and the phase history, $\varphi(t)$ (Rosen et al., 2000).

Only the principal values of the phase, modulo 2π, can be measured from the complex-valued resolution element. The total range difference between the two observation points that the phase represents in general can be many multiples of the radar wavelength or, expressed in terms of phase, many multiples of 2π. The typical approach for determining the unique phase that is directly proportional to the range difference is to first determine the relative phase between pixels via the so-called 'phase-unwrapping' process. This connected phase field will then be adjusted by an overall constant multiple of 2π. The second step determines this required multiple of 2π and is referred to as 'absolute phase determination.' Figure 8 shows the principal value of the phase, ϕ_m, the unwrapped phase, ϕ_{unw}, and absolute phase, ϕ_{top}, for a pixel.

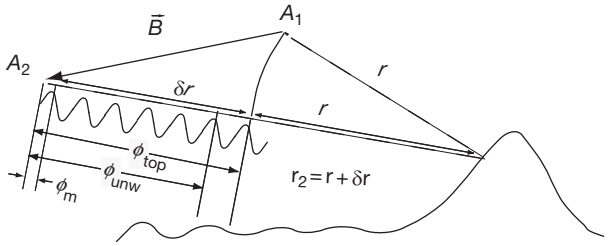


Figure 8 Phase in interferogram depicted as cycles of electromagnetic wave propagating a differential distance δr. Phase in the interferogram is initially known modulo 2π: $\phi_m = W(\phi_{top})$, where ϕ_{top} is the topographically induced phase and $W()$ is an operator that wraps phase values into the range $-\pi < \phi \leq \pi$. After unwrapping, relative phase measurements between all pixels in the interferogram are determined up to a constant multiple of 2π: $\phi_{unw} = \phi_m + 2\pi k_{unw}$, where k_{unw} is a spatially variable integer dependent on the pixel coordinates of the interferogram. Absolute phase determination is the process to determine the overall multiple of $2\pi k_{abs}$ that must be added to the phase measurements so that it is proportional to the range difference. The reconstructed phase is then $\phi_{top} = \phi_m + 2\pi k_{unw} + 2\pi k_{abs}$.

3.12.2.2 Interferometric Baseline and Height Reconstruction

In order to generate topographic maps or data for other geophysical applications using radar interferometry, we must relate the interferometric phase and other known or measurable parameters to the topographic height. It is also desirable to derive the sensitivity of the interferometrically determined topographic measurements to the interferometric phase and other known parameters. In addition, interferometric observations have certain geometric constraints that preclude valid observations for all possible image geometries.

The interferometric phase as previously defined is proportional to the range difference from two antenna locations to a point on the surface. This range difference can be expressed in terms of the vector separating the two antenna locations, called the interferometric baseline. The range and azimuth position of the sensor associated with imaging a given scatterer depends on the portion of the synthetic aperture used to process the image (see Appendix A.1). Therefore, the interferometric baseline depends on the processing parameters and is defined as the difference between the locations of the two antenna phase center vectors at the time when a given scatterer is imaged.

The equation relating the scatterer position vector, $\vec{T}$, a reference position for the platform $\vec{P}$, and the look vector, $\vec{l}$, is

$$\begin{aligned}\vec{T} &= \vec{P} + \vec{l} \\ &= \vec{P} + r\hat{l}\end{aligned} \quad [3]$$

where r is the range to the scatterer and $\hat{l}$ is the unit vector in the direction of $\vec{l}$ (Figure 9). The position $\vec{P}$ can be chosen arbitrarily but is usually taken as the position of one of the interferometer antennas. Interferometric height reconstruction is the determination of a target's position vector from known platform ephemeris information, baseline information, and the interferometric phase. Assuming $\vec{P}$ and r are known, interferometric height reconstruction amounts to the determination of the unit vector $\hat{l}$ from the interferometric phase. Letting $\vec{B}$ denote the baseline vector from antenna 1 to antenna 2, setting

$\vec{P}=\vec{P}_1$ and defining $\vec{B}=\vec{P}_2-\vec{P}_1$ and $B=\left|\vec{B}\right|\equiv\left\langle\vec{B},\vec{B}\right\rangle^{1/2}$, we have the expression for the interferometric phase

$$\begin{aligned}\phi&=\frac{2\pi p}{\lambda}(r_2-r_1)\\&=\frac{2\pi p}{\lambda}\left(\left|\vec{l}_2\right|-\left|\vec{l}_1\right|\right)\\&=\frac{2\pi p}{\lambda}r_1\left[\left(1-\frac{2\left\langle\hat{l}_1,\vec{B}\right\rangle}{r_1}+\left(\frac{B}{r_1}\right)^2\right)^{1/2}-1\right]\end{aligned}\quad[4]$$

where $p=2$ for repeat-track systems and $p=1$ for two-aperture systems with a single transmitter and two receivers (Rosen et al., 2000) and the subscripts refer to the antenna number. This expression can be simplified assuming $B\ll r$ by Taylor-expanding eqn [4] to first order to give

$$\phi\approx-\frac{2\pi p}{\lambda}\left\langle\hat{l}_1,\vec{B}\right\rangle\quad[5]$$

illustrating that the phase is approximately proportional to the projection of the baseline vector on the look direction (Zebker and Goldstein, 1986).

When the baseline lies entirely in the plane of the look vector and the nadir direction, we have $\vec{B}=(B\cos(\alpha),B\sin(\alpha))$,

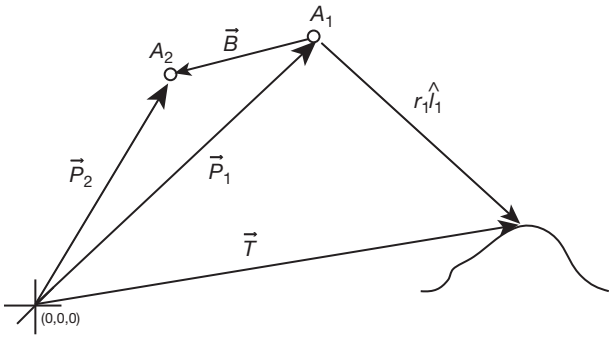


Figure 9 Vectors describing the relationship between the phase centers of the radar antennas defining the interferometer and the surface location, as described in the text.

where α is the angle the baseline makes with respect to a reference horizontal plane. Then, eqn [5] can be rewritten as

$$\phi=-\frac{2\pi p}{\lambda}B\sin(\theta-\alpha)\quad[6]$$

where θ is the look angle, the angle the line-of-sight vector makes with respect to nadir, shown in Figure 10. The intrinsic fringe frequency in the slant plane interferogram is given as

$$\begin{aligned}&\frac{\partial\phi}{\partial r}\\&=\frac{2\pi p}{\lambda}B\cos(\theta-\alpha)\frac{1}{r\sin\theta}\left[-\frac{r}{h_p+r_e}+\cos\theta+\frac{h_0+r_e}{h_p+r_e}\frac{\sin\tau_c}{\sin(i-\tau_c)}\right]\end{aligned}\quad[7]$$

where

$$\sin i=\frac{h_p+r_e}{h_0+r_e}\sin\theta\quad[8]$$

and i is the local incidence angle relative to a spherical surface, h_p is the height of the platform, and τ_c is the surface slope angle in the cross track direction as defined in Figure 10 at the left. From eqn [7], the fringe frequency is proportional to the perpendicular component of the baseline. As $B_\perp$ increases or as the local terrain slope approaches the look angle, the fringe frequency increases. Also from eqn [7], the fringe frequency is inversely proportional to λ, thus longer wavelengths result in lower fringe frequencies. If the phase changes by 2π or more across the range resolution element, Δr, the different contributions within the resolution cell do not add to a well-defined phase resulting in what is commonly referred to as the decorrelation of the interferometric signal. Thus, in interferometry, an important parameter is the critical baseline, defined as the perpendicular baseline at which the phase rate reaches 2π per range resolution element. From eqn [7], the critical baseline satisfies the proportionality relationship

$$B_{\perp,\mathrm{crit}}\propto\frac{\lambda}{\Delta r}\quad(9)$$

This is a fundamental constraint for interferometric radar applied to natural (distributed scattering) surfaces. Point

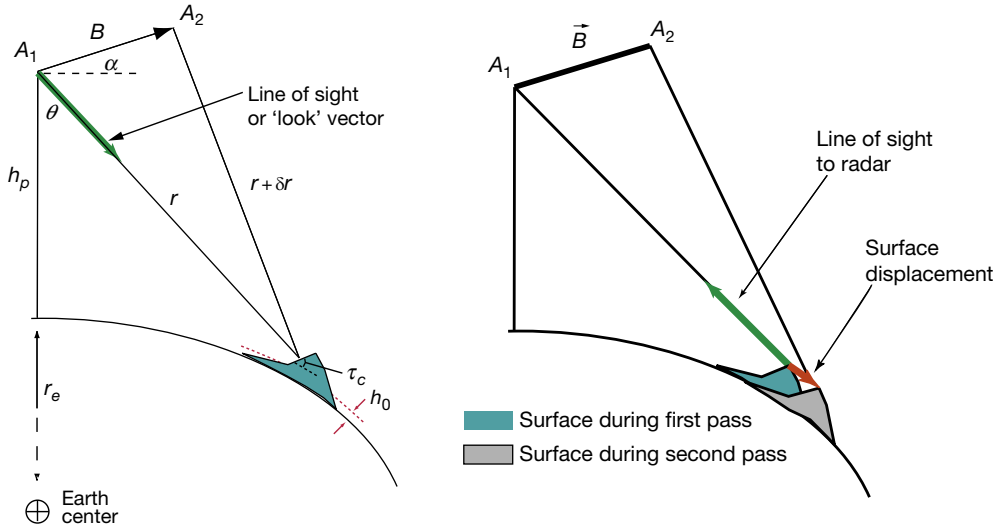


Figure 10 SAR interferometry imaging geometry in the plane normal to the flight direction for topography (left) and deformation (right) mapping.

targets, sometimes called permanent scatterers, can maintain phase coherence beyond this critical baseline however. Difficulty in phase unwrapping increases (see Section 3.12.2.4) as the fringe frequency approaches this critical value.

The fringe variation in the interferogram is 'flattened' by subtracting the expected phase from a surface of constant elevation. The resulting fringes follow the natural topography more closely. Letting $\hat{l}_0$ be a unit vector pointing to a surface of constant elevation, h_0, the flattened phase, ϕ_{flat}, is given by

$$\phi_{\text{flat}} = -\frac{2\pi p}{\lambda}\left(\left\langle \hat{l}, \vec{B}\right\rangle - \left\langle \hat{l}_0, \vec{B}\right\rangle\right) \quad [10]$$

where

$$\hat{l}_0 = (\sin\theta_0 - \cos\theta_0) \quad [11]$$

and $\cos\theta_0$ is given by the law of cosines

$$\cos\theta_0 = \frac{r_0^2 + (r_e + h_p)^2 - (r_e + h_0)^2}{2(r_e + h_p)r_0} \quad (12)$$

assuming a spherical Earth with radius r_e and a slant range to the reference surface r_0.

Equation [10] can be simplified by expanding the look angle $\theta = \theta_0 + \delta\theta$, where $\delta\theta$ is the contribution to the look angle at range r_0 from the topographic relief relative to a reference surface, and θ_0 is the look angle to the reference surface at range r_0. If the topographic relief is represented by δz, then $\delta\theta = \delta z / r_0 \sin\theta_0$, and

$$\begin{aligned}\phi_{\text{flat}} &= -\frac{2\pi p}{\lambda}\left(\left\langle \hat{l}, \vec{B}\right\rangle - \left\langle \hat{l}_0, \vec{B}\right\rangle\right) \\ &\approx -\frac{2\pi p}{\lambda} B_\perp \frac{\delta z}{r_0 \sin\theta_0}\end{aligned} \quad [13]$$

is the component of baseline perpendicular to the line of sight. Equation [13] tells us that

- flattened fringes are proportional to the topographic height and thus a poor man's topographic map can be generated by flattening the phase and examining the fringes;
- flattened fringes are proportional to the perpendicular component of the baseline; (For zero baseline, there are no fringes, even if there is a large parallel component of the baseline. For large baseline, there are many cycles of phase change for a given topographic change. From this equation, we can define h_a, the *ambiguity height*, as $h_a = \partial h / \partial\phi = \lambda r_0 \sin\theta_0 / 2\pi p B_\perp$.)
- in the absence of topographic variations, there is still an intrinsic variation of fringes across an interferogram given by the flattening phase.

3.12.2.3 Differential Interferometry

The theory just described assumes that the imaged surface is stationary over time or that the surface is imaged by the interferometer at a single instant. When there is motion of the surface between radar observations, there is an additional contribution to the interferometric phase variation. Figure 10 at right shows the geometry when a surface displacement occurs between the observation at $\vec{P}_1$ (at time t_1) and the observation at $\vec{P}_2$ (at $t_2 > t_1$). In this case, $\vec{l}_2$ becomes

$$\begin{aligned}\vec{l}_2 &= \vec{T} + \vec{D} - \vec{P}_2 \\ &= \vec{l}_1 + \vec{D} - \vec{B}\end{aligned} \quad [14]$$

where $\vec{D}$ is the displacement vector of the surface from t_1 to t_2. The interferometric phase expressed in terms of this new vector is

$$\phi = \frac{4\pi}{\lambda}\left(\left\langle \vec{l}_1 + \vec{D} - \vec{B}, \vec{l}_1 + \vec{D} - \vec{B}\right\rangle^{1/2} - r_1\right) \quad [15]$$

Assuming as in the preceding text that $\left|\vec{B}\right|$, $\left|\vec{D}\right|$, and $\left|\left\langle \vec{B}, \vec{D}\right\rangle\right|$ are all much smaller than r_1, the phase reduces to

$$\phi = \frac{4\pi}{\lambda}\left(-\left\langle \vec{l}_1, \vec{B}\right\rangle + \left\langle \vec{l}_1, \vec{D}\right\rangle\right) \quad [16]$$

Typically, for spaceborne geometries $B < 1$ km, and D is of order meters, while $r_1 \approx 600$–800 km. These values justify the usual formulation in the literature that

$$\phi_{\text{obs}} = \phi_{\text{topography}} + \phi_{\text{displacement}} \quad [17]$$

In some applications, the displacement phase represents a nearly instantaneous translation of the surface resolution elements, for example, earthquake deformation. In other cases, such as glacier motion, the displacement phase represents a motion tracked over the time between observations. Intermediate cases include slow and/or variable surface motions, such as volcanic inflation or surging glaciers. Equations [16] and [17] highlight that the interferometer measures the projection of the displacement vector in the radar line-of-sight (LOS) direction. To reconstruct the vector displacement, observations must be made from different LOS directions (see Section 3.12.4.7).

The topographic phase term is not of interest for displacement mapping, and must be removed. Several techniques have been developed to do this. They all essentially derive the topographic phase from another data source, either a DEM or another set of interferometric data. The selection of a particular method for topography measurement depends heavily on the nature of the motion (steady or episodic), the imaging geometry (baselines and time separations), and the availability of data.

We emphasize the increase in precision of the interferometric displacement measurement relative to topographic mapping precision. Consider a discrete displacement event such as an earthquake where the surface moves by a fixed amount $\vec{D}$ in a short time period. Neither a pair of observations acquired before the event (pair 'a') nor a pair after the event (pair 'b') would measure the displacement directly but together would measure it through the change in topography. According to eqn [15], and assuming the same imaging geometry for 'a' and 'b' without loss of generality, the phase difference between these two interferograms (that is the difference of phase differences) is

$$\begin{aligned}\phi = \phi_a - \phi_b = \frac{4\pi}{\lambda}\Bigg[&\left(\left\langle \vec{l}_1 - \vec{B}, \vec{l}_1 - \vec{B}\right\rangle^{1/2} - r_1\right) \\ &- \left(\left\langle \vec{l}_1 + \vec{D} - \vec{B}, \vec{l}_1 + \vec{D} - \vec{B}\right\rangle^{1/2} - \left\langle \vec{l}_1 + \vec{D}, \vec{l}_1 + \vec{D}\right\rangle^{1/2}\right)\Bigg] \\ &= 0\end{aligned} \quad [18]$$

to first order, because $\vec{D}$ appears in both the expression for $\vec{l}_2$ and $\vec{l}_1$. The nature of the sensitivity difference inherent

between eqns [16] and [18] can be seen in the 'flattened' phase (see eqn [13]) of an interferogram, often written (Rosen et al., 1996)

$$\phi = -\frac{4\pi}{\lambda} B \cos(\theta_0 - \alpha) \frac{z}{r_0 \sin\theta_0} + \frac{4\pi}{\lambda} \delta r_{\text{disp}} \qquad [19]$$

where δr is the surface displacement between imaging times in the LOS direction, and z is the topographic height above the reference surface. In this formulation, the phase difference is far more sensitive to *changes* in topography (surface displacement) than to the topography itself. From eqn [19], $\delta r = \lambda/2$ gives one cycle of phase difference, while z must change by a substantial amount, essentially r_0/B, to affect the same phase change. For example, for ERS, $\lambda = 5.6$ cm, $r_1 \approx 830$ km, and typically $B \leq 200$ m, implying $\delta r_{\text{disp}} = 2.8$ cm to generate one cycle of phase, $z \geq 450$ cm to have the same effect.

The time interval over which displacement is measured must be matched to the geophysical signal of interest. To capture an evolving magmatic event, the temporal baseline may need to be of the order of days or less. At the other extreme, temporal baselines of several years may be required to make accurate measurements of slow deformation processes such as interseismic strain.

Reconstruction of the scatterer position vector depends on the knowledge of the platform location, the interferometric baseline length, the orientation angle, and the interferometric phase. To generate accurate topographic or displacement maps, radar interferometry places stringent requirements on knowledge of the platform and baseline vectors, as well as the intrinsic accuracy of the phase measurements and, in the case of differential interferometry, supporting topographic data sets. One source of phase noise is the refractivity of the atmosphere, which varies along the radar propagation path in space and time. Refractivity fluctuation due to turbulence in the atmosphere is a minor effect for two-aperture cross track interferometers but is much more important for repeat-track systems (Rosen et al., 1996). Sensitivities to these parameters are discussed in detail in Rosen et al. (2000) and Zebker et al. (1994).

3.12.2.4 Phase Unwrapping

While methods have been developed for geophysical interpretation of deformation observations using wrapped interferometric phase (Feigl and Thurber, 2009; Price and Sandwell, 1998), most current applications still rely on unwrapping the phase to remove the modulo-2π ambiguity before estimating topography or surface displacement. A simple approach to phase unwrapping would be to form the first differences of the phase at each image point in either image dimension as an approximation to the derivative and then integrate the result. Direct application of this approach, however, allows local errors due to phase noise to propagate, causing errors across the full SAR scene. The unwrapped solution should, within a constant of integration, be independent of the path of integration such that in the error-free case, the integral of the differenced phase about a closed path is zero. Phase inconsistencies are therefore indicated by nonzero results when the phase difference is summed around the closed paths formed by each mutually neighboring set of four pixels. These points have either a positive or negative integral (by convention performed in clockwise paths). Integration of the differenced phase about any closed path yields a value equal to the sum of the enclosed points of inconsistency. As a result, paths of integration that encircle a nonzero sum must be avoided. In the so-called 'branch-cut' methods (Goldstein et al., 1988), this is accomplished by connecting oppositely signed points of phase inconsistency with lines that the path of integration cannot cross. Once these barriers have been selected, phase unwrapping is completed by integrating the differenced phase subject to the rule that paths of integration do not cross the barriers.

The phase-unwrapping problem becomes particularly difficult when the phase in the interferogram is intrinsically discontinuous or highly decorrelated, due to layover problems or true shear topography. Most algorithms are based on the assumption that the phases are continuous, and often natural phase discontinuities, often corrupted with inherent phase noise, are difficult to interpret.

A full treatment of phase unwrapping for geodetic imaging applications is beyond the scope of this chapter. The literature describing unwrapping approaches is quite large (Rosen et al., 2000), with initial development of branch-cut techniques for InSAR applications followed by over a decade of exploration of other techniques. The most widely applied unwrapping algorithms, including branch-cut and statistical cost network flow techniques (Chen and Zebker, 2001), yield unwrapped phase images that are multiples of 2π of the original wrapped phase image. In the case of branch-cut algorithms, there are often regions that are blocked off from unwrapping by the barriers that form a complete circuit. For network flow, the entire image is unwrapped. In all cases, there will be individual pixels or areas that are placed on the wrong multiple of 2π, and it is often quite difficult to identify these points without additional information. If multiple interferograms with a variety of baselines are available, it can be advantageous to compare the unwrapped solutions for each to determine consistency of the solution or use the inherent redundancy to jointly estimate a unique solution for the underlying signal (Fornaro et al., 2006).

3.12.2.5 Correlation

The relationship between the scattered electromagnetic fields seen at the interferometric receivers after image formation is characterized by the complex correlation function, γ, defined as

$$\gamma = \frac{\langle \Gamma_1 \Gamma_2^* \rangle}{\sqrt{\langle |\Gamma_1|^2 \rangle \langle |\Gamma_2|^2 \rangle}} \qquad [20]$$

where Γ_i represents the SAR reflectivity at the i antenna and angular brackets denote averaging over the ensemble of speckle realizations. For completely coherent scatterers such as point scatterers, we have that $\gamma = 1$, while $\gamma = 0$ when the scattered fields at the antennas are independent. The magnitude of the correlation $|\gamma|$ is often referred to as the 'coherence.' (Several authors distinguish between the 'coherence' properties of fields and the correlation functions that characterize them, e.g., Born and Wolf, 1989, whereas others do not make a distinction.)

In general, correlation is controlled by a number of effects:

$$\gamma = \gamma_N \gamma_G \gamma_Z \gamma_T \quad [21]$$

where γ_N is the correlation influenced by noise in the radar system and processing approach; γ_G is that influenced by the different observing geometries; γ_Z describes the influence on correlation of the vertical extent of scatterers, for example, due to vegetation; and γ_T describes the influence of repositioning of scatterers within a resolution element over time (Bamler and Hartl, 1998; Li and Goldstein, 1990; Rodriguez and Martin, 1992; Rosen et al., 2000; Zebker and Villasenor, 1992). It is often more convenient to discuss decorrelation, defined as $\delta_X = 1 - \gamma_X$, where X is N, G, Z, or T.

Geometric decorrelation, δ_G, also called baseline or speckle decorrelation, is due to the fact that, after removing the phase contribution from the center of the resolution cell, the phases from the scatterers located away from the center are slightly different at each antenna (see Figure 7). The degree of decorrelation can then be estimated from the differential phase of two points located at the edges of the area obtained by projecting the resolution cell phase from each scatterer within the resolution cell, as shown in Figure 7. Using this simple model, one can estimate that the null-to-null angular width of the correlation function, $\Delta\theta$, is given by

$$\Delta\theta \approx \frac{B_\perp}{r} \approx \frac{\lambda}{\Delta r_l} \quad [22]$$

where $B_\perp$ is the projection of the interferometric baseline onto the direction perpendicular to the look direction, and Δr_l is the projection of the ground resolution cell along the same direction, as illustrated in Figure 11.

The geometric correlation term is present for all scattering situations and depends on the system parameters and the observation geometry.

Under the right conditions, one can adjust the processing approach to create the local condition of no geometric decorrelation (Gatelli et al., 1994). This is accomplished by noting that for the specific geometry depicted in the lower right of Figure 7, the different projected planes of the two observations amount to a small linear phase gradient difference between the two observations, which is equivalent to a frequency shift. In other words, the geometric decorrelation term in principle can be eliminated by proper choice of center frequencies for two observations. In practice, however, the relative observation geometry depends on the look angle and surface slope, so that adaptive iterative processing is required in order to implement the approach exactly (Davidson and Bamler, 1999).

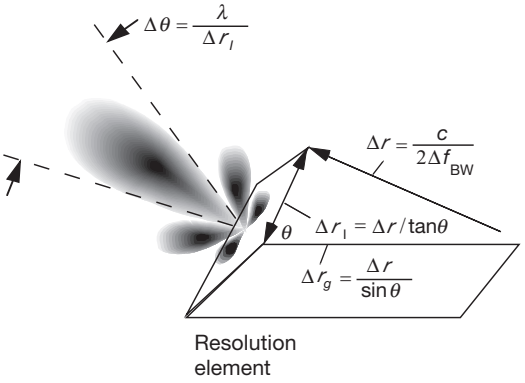

$\Delta\theta = \frac{\lambda}{\Delta r_l}$
$\Delta r = \frac{c}{2\Delta f_{BW}}$
θ
$\Delta r_l = \Delta r / \tan\theta$
$\Delta r_g = \frac{\Delta r}{\sin\theta}$
Resolution element

Figure 11 A view of geometric decorrelation showing the effective beam pattern of a ground resolution element 'radiating' to space. The mutual coherence field propagates with radiation beamwidth in elevation of $\Delta\theta \approx \lambda/\Delta r_l$. These quantities are defined in the figure.

The volumetric correlation, γ_Z, can be understood in terms of an effective increase in the size of the projected range cell Δr_l (Rosen et al., 2000). When the scattering elements in a given range cell are extended vertically, such as when trees or building rise up from the surface, the contributions to the net backscatter phase from a given pixel are extended over a larger projected area. Imagine in Figure 11 that a tall structure existed within the pixel and that the brightest scattering point was at the top. The energy from that point scattered back toward the radar intersects the reference surface above the rectangle shown. Thus, these vertically extended scatterers extend the effective size of the projected range cell and consequently decrease the correlation in that pixel according to eqn [22]. If the range resolution is infinitesimally small, then the volume decorrelation effect can be understood as being due to the geometric decorrelation from a plane cutting through the scattering volume perpendicular to the look direction.

In repeat-pass systems, temporal decorrelation, δ_T, occurs when the surface changes between the times when the images forming an interferogram are acquired (Zebker and Villasenor, 1992). As scatterers become randomly rearranged over time, the detailed speckle patterns of the image resolution elements differ from one image to the other, so the images no longer correlate.

While it is difficult to describe these effects analytically, this can often be a strong limitation on the accuracy of repeat-pass data, so a few illustrative examples are in order. Open water where the surface is roughened by wind or turbulence is constantly changing over time, so two images will completely decorrelate ($\gamma = 0$). Similarly, an agricultural field where the entire surface has been turned over due to tilling will have $\gamma = 0$. Standing water with no vegetation present above water will also completely decorrelate because no signal is scattered back toward the radar. However, with vegetation present above the surface, the water serves as a mirror that permits signal return from the water as scattered off the vegetation (Alsdorf et al., 2001; Wdowinski et al., 2004). Rain can change the arrangement of vegetation on a surface (e.g., sagging branches or stalks) reducing the correlation by an amount dependent on the density of altered vegetation. In some cases, as the surface dries, the vegetation bounces back to its original position and correlation is at least partially restored. However, wind typically will alter the positions of scatterers in vegetation canopies over time, so correlation is generally degraded in vegetation. Snow can destroy correlation in the winter months, with correlation restored after the snow is gone. For interferometry applied to geophysical processes, we rely on the block motion of pixels without scatterer rearrangement to provide estimates of the geodetic motion. These effects and changes due to ground shaking (building collapse, landslides, liquefaction, etc.) can impair the ability to measure displacements. On the other hand, it can also be a means for understanding the nature of the surface and the severity of the geophysical effects on the ground.

In addition to these field correlations, thermal noise in the interferometric channels also introduces phase noise in the interferometric measurement. The correlation due to thermal noise alone can be written as

$$\gamma_N = \frac{1}{\sqrt{1+\mathrm{SNR}_1^{-1}}\sqrt{1+\mathrm{SNR}_2^{-1}}} \quad [23]$$

where SNR_i denotes the signal-to-noise ratio for the i channel (Zebker and Villasenor, 1992). In addition to thermal noise, which is additive, SAR returns also have other noise components, due to, for example, range and Doppler ambiguities. An expression for the decorrelation due to this source of error can only be obtained for homogeneous scenes, since, in general, the noise contribution is scene-dependent. Typically, for simplicity, these ambiguities are treated as additive noise as part of the overall system noise floor.

The effect of decorrelation is the apparent increase in noise of the estimated interferometric phase. Rodríguez and Martin (1992) presented the analytic expression for the Cramer–Rao bound (Sorenson, 1980) on the phase variance

$$\sigma_\phi^2 = \frac{1}{2N_L}\frac{1-\gamma^2}{\gamma^2} \quad [24]$$

where N_L is the number of independent samples used to derive the phase and is usually referred to as the 'number of looks.' The actual phase variance approaches the limit equation [24] as the number of looks increases and is a reasonable approximation when the number of looks is greater than 4. An exact expression for the phase variance can be obtained starting from the probability density function for the phase when $N_L = 1$ and then extended for an arbitrary number of looks (Goodman, 1985; Joughin et al., 1994; Lee et al., 1992; Touzi and Lopes, 1996). The expressions, however, are quite complicated and must be evaluated numerically in practice.

Note that the estimate of the correlation is usually accomplished by computing the expectation operations in eqn [20] as spatial averages over a number of pixels in an interferometric pair. This leads to biased estimates of the correlation, and care must be exercised in interpreting the estimate. For example, in open water, where the actual coherence of the fields is zero, the correlation estimate will produce decidedly nonzero estimates, in the range of 0.1–0.3, depending on the number of samples used in the estimate, N_L. The estimator is a random variable with a probability distribution shape that depends on the intrinsic coherence and the number of samples used in the estimate. In the limit, where only one sample is used, the correlation estimate will be 1!

3.12.2.6 Processing Data

There exist both commercial and freely available software for conventional InSAR processing (e.g., Rosen et al., 2004; Sandwell et al., 2011). While they may differ in detail, they must all follow a basic processing flow. Figure 12 presents such a flow, derived from the authors experience developing the ROI_PAC (Repeat Orbit Interferometry Package) software suite (Rosen et al., 2004). This flow diagram explicitly calls out potential iterative cycles and the use of external data and intermediary models. All major phases of this processing are described in the text. Given quality data and metadata, an initial complete processing from raw data to a georeferenced deformation image can now be done automatically and quickly (in under a few hours) on an average laptop computer. Subsequent iterative refinements to maximize the quality and quantity of the observations can require significantly more effort. Hooper et al. (2012) in an appendix provided a nonexhaustive list of popular interferometry software packages, as well as a table of spaceborne sensors that have interferometrically viable data. They note that the field is rapidly expanding, and one common means for new researchers to deeply understand the methods of interferometry is to write their own software or modify others. Particularly for the freely distributed software packages such as ROI_PAC mentioned in the preceding text, a number of variants are in use throughout the research community.

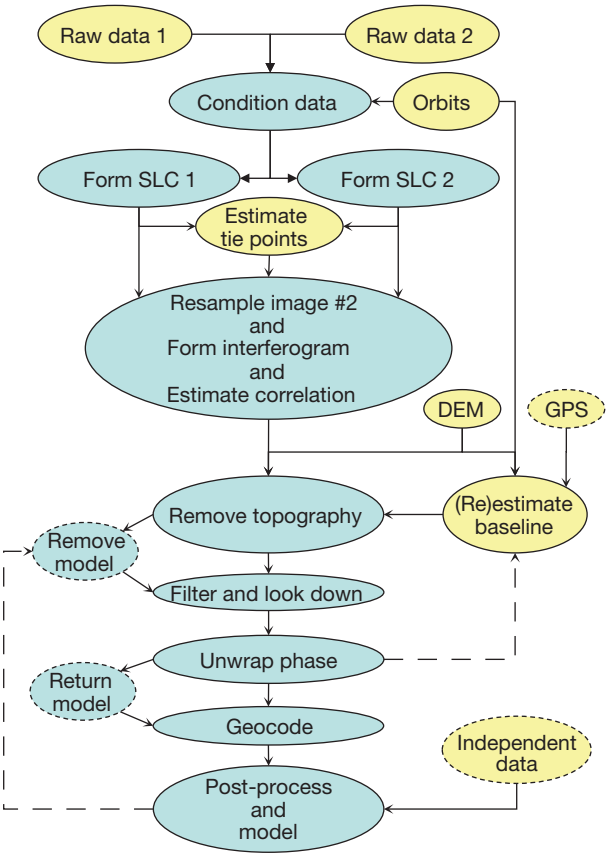


Figure 12 Representative differential InSAR processing flow diagram. Blue bubbles represent image output; yellow ellipses represent nonimage data. Flow is generally down the solid paths, with optional dashed paths indicated potential iteration steps. The following acronyms are used: DEM, digital elevation model; SLC, single-look complex image.

3.12.3 InSAR-Related Techniques

3.12.3.1 Wide-Swath Interferometry

For previously flown SAR systems operating in traditional 'strip map' mode, where a continuous stream of pulses is transmitted and received to allow formation of an arbitrarily long strip

image, the width of the strip's swath has been limited to somewhat less than 100 km. As discussed in Section 3.12.5, SAR antennas must satisfy particular minimum area criteria to ensure noise due to ambiguities below a required level. For wide swath, they must also be quite long, which can be difficult and costly to implement. For example, to achieve a 300 km swath in typical Earth orbits using the typical strip mapping method, the antenna would have to be over 40 m in length. To achieve wide swaths with an antenna sized for a narrow swath strip imaging, ScanSAR techniques were developed (Moore et al., 1981; Tomiyasu, 1981). ScanSAR requires an antenna with rapid electronic steering capability in the elevation direction. In ScanSAR, the antenna is electronically steered to a particular elevation angle and radar pulses are transmitted and echoes received for a time period that is a fraction (say one-tenth) of the synthetic aperture time. After that 'dwell period,' also known as the 'burst period,' the antenna is electronically steered to another elevation angle (and other radar parameters such as the PRF, bandwidth, and antenna beam shape are changed), and observations are made for another short dwell period. This process is repeated until each of the elevation directions, needed to observe the entire wide swath, is obtained, at which point the entire cycle of elevation dwell periods is repeated. This is illustrated in Figure 13.

For any given illuminated range of elevation angles, or subswath, there are large gaps in the receive echo timeline, yet after processing the data, a continuous image can be formed. This is true because the extent of the antenna beam in the along-track direction on the ground is proportional to the synthetic aperture length. As long as the dwell periods for any given subswath occur more than once in the synthetic aperture time, there is guaranteed continuous coverage of all points on the ground.

The ScanSAR method places additional constraints on its use for InSAR related to the truncation of the observing time as the beam is switched from one location to another from burst to burst. Each pulse in a ScanSAR burst contains the full Doppler spectrum of information (see Doppler discussion in the appendix). However, any given scattering element within a burst period only contributes a portion of its full Doppler history because it is not observed over the full synthetic aperture time (Bamler and Eineder, 1996; Guarnieri and Prati, 1996; Lanari et al., 1998). Thus, each scattering element is only resolved commensurate with the burst period relative to the synthetic aperture time: if the resolution in strip mode is $L/2$, then the resolution in ScanSAR mode is $(L/2) \times T_a/T_b$, where T_a is the synthetic aperture time, and T_b is the burst duration. If one were to attempt to achieve maximal resolution possible, one would divide the synthetic aperture time by the number of subswaths needed and set the burst duration to this time. However, as described following, it is generally better to create several short bursts within one synthetic aperture. This degrades resolution further but improves the radiometric characteristics of the data.

In interferometry, a scattering element must be observed with the same range of Doppler frequencies from one pass to

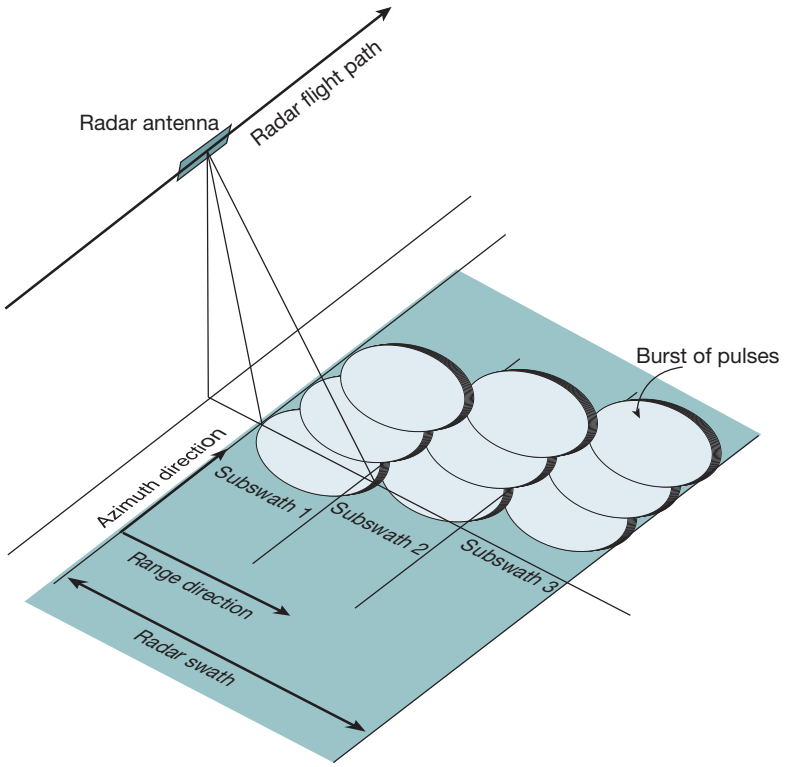


Figure 13 ScanSAR for a 3-beam system. The radar transmits a burst of pulses to illuminate subswath 1 and then electronically switches to point to subswath 2 and then subswath 3. This cycle is repeated for the extent of the data taken. The aspect ratio of the beams in this figure is highly stretched in azimuth to illustrate the pulsed behavior for the bursts and the significant beam overlap from burst to burst within a subswath. This overlap allows for construction of continuous maps.

another. Only then will the images have similar speckle patterns and therefore be coherent. This implies that from pass to pass, each observation must observe from the same group of angles. In the case of strip mode SAR, this implies that the intrinsic pointing of the antenna beam be the same from pass to pass, ensuring the Doppler history of each scattering element will follow the same course. For ScanSAR, this condition also implies a timing constraint on the bursts. Each burst must occur at the same location in space relative to a scattering element from pass to pass. The Shuttle Radar Topography Mission employed ScanSAR techniques to map the Earth's topography in a single 10-day shuttle mission (Farr et al., 2007). With two antennas separated by a 60-m boom, and timing uniformly controlled throughout the system, it was possible to easily synchronize bursts for single-pass interferometry. For repeat-pass systems, with drift over time of the time reference of the radar, and orbits that are not always known to adequate accuracy in advance of an observation, the ScanSAR burst timing constraint is challenging to meet, particularly with short bursts. The potential benefits of wide-swath InSAR mapping have nonetheless been recognized through a number of successful repeat-pass demonstrations using Radarsat, Envisat, and ALOS-1 (e.g., Holzner and Bamler, 2002). Radar satellites scheduled to be operational in the 2013–14 time frame such as ALOS-2 and Sentinel-1 have been designed specifically to synchronize bursts from orbit to orbit over the life of the mission. For Sentinel-1 (Snoeij et al., 2008), the primary observation mode is planned to be a ScanSAR-like mode called TOPS, short for 'Terrain Observation by Progressive Scans,' in which the radar beam is actively swept over a range of azimuth angles around broadside from backward to forward over the duration of each burst. This mode allows for longer bursts, which relaxes timing constraints, and improves the radiometric characteristics of the data (De Zan and Monti Guarnieri, 2006).

Figure 14 illustrates the workflow associated with ScanSAR interferometry. The principal difference between this workflow and the conventional strip map approach is that the interferograms are formed at the burst level before combining to form a large-area interferogram. This is necessary in general because the bursts do not overlap precisely, and filtering tailored to the misalignment is needed to optimize coherence (Guarnieri and Prati, 1996).

ScanSAR can be used to densify a deformation time series. For a satellite in a long repeat-period orbit, for example, 32 days, the ground swath will be sized to achieve global coverage, around 85 km for the 32-day repeat-period orbit. So from an interferometric point of view, using ScanSAR with this period does not improve on the interferometric interval intrinsically. However, ScanSAR will increase the number of times a given scattering element will be observed by a factor of the number of ScanSAR beams. For a 4-beam ScanSAR with

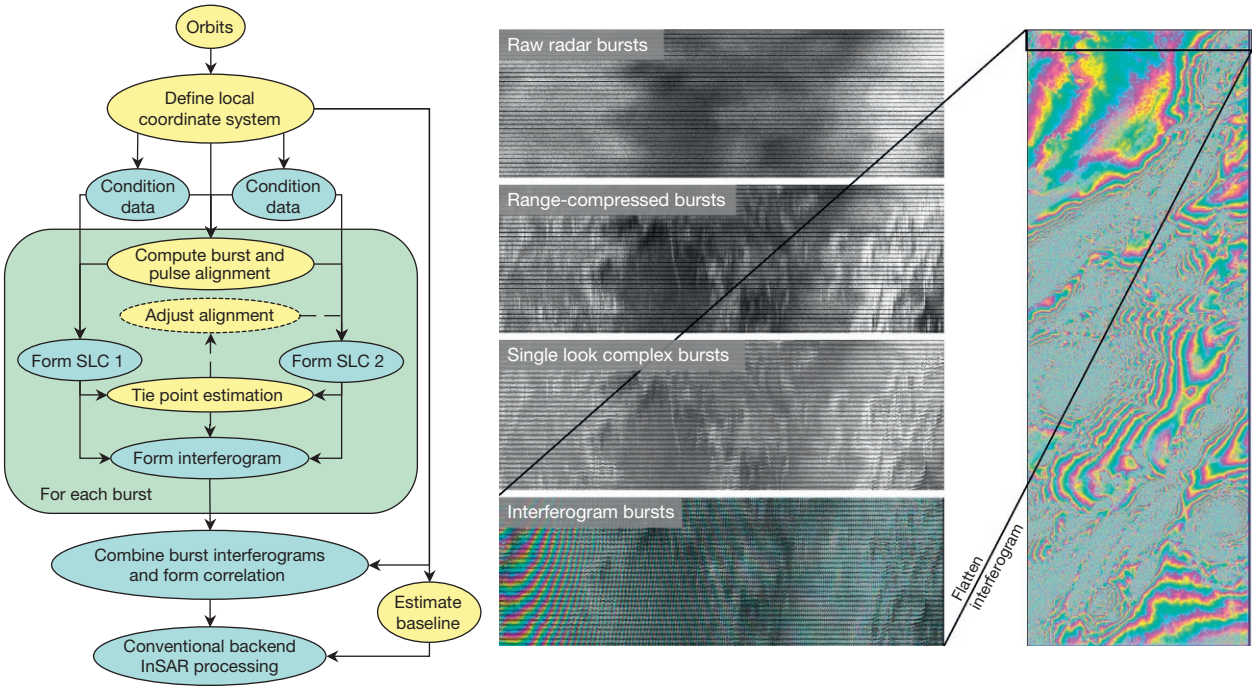


Figure 14 (Left) Representative differential InSAR processing flow diagram for ScanSAR interferometry. Blue bubbles represent image output; yellow ellipses represent nonimage data. Flow is generally down the solid paths, with optional dashed paths indicated potential iteration steps. (Right) Illustration of the workflow by explicit example. The top three black and white panels show a collection of bursts from one scene for one ScanSAR subswath. The top panel displays raw radar data, with a small artificial gap between bursts to delimit them. The next panel shows the bursts after range compression – features appear to be sharper in the range (across image) dimension. The next panel shows the bursts after azimuth compression – single-look complex bursts – features are now sharp in both dimensions, and the existence of surface features in three or four successive bursts is apparent, showing the intrinsic overlap of the data. The bottom center panel shows a burst interferograms formed from two SLC burst sequences. After combination and flattening, the bursts form a continuous interferogram as shown on the right, with the left sequence contained in the box at the top as indicated. This interferogram, as well as those of the other subswaths, can then be processed as usual.

340 km swath in a 32-day repeat orbit, a scattering element will be observed roughly every 8 days, so one could make interleaved 32-day repeat interferograms. Alternatively, one could place the radar in a shorter repeat-period orbit, for example, 8 days, and observe always in ScanSAR mode with four beams. This would allow 8-day interferograms with no interleaving. The advantage of the latter is that decorrelation will be less of an issue in the basic repeat-pass measurement. The advantage of the former is that there will be greater angular diversity in the measurement, potentially better constraining models of deformation.

It is also possible to combine ScanSAR data with strip mode data where the observation subswath of the ScanSAR image coincides with the strip mode swath (Monti Guarnieri and Prati, 1996). Since the strip mode data are continuously pulsed with no gaps, there are no burst alignment issues; in processing, one must simply choose the strip mode pulses that align with the ScanSAR bursts. Of course, the strip mode data only cover a portion of the complete ScanSAR swath, so this method is most beneficial when ScanSAR data are available to fill in observational gaps in a strip mode deformation time series (Pepe et al., 2011).

3.12.3.2 Permanent Scatterers and Time-Series Analysis

The two major error sources in InSAR are decorrelation due to temporal and geometric effects and phase errors introduced by the spatially and temporally random variations of the refractive index of the atmosphere and ionosphere. Decorrelation creates areas that are spatially disjoint, and irregularly so over time, leading to difficulty in interpreting the geodetic measurements. Because the repeat period of the orbital satellites is on the order of weeks, temporal sampling of dynamic phenomena is also poor. Continuous GPS measurements share similar characteristics – spatially disjoint measurements that also are subject to atmospheric effects – but have the advantage of dual-frequency measurements for ionospheric correction and continuous temporal sampling while observing multiple satellites, which allows for a fine sampling of atmospheric variability leading to robust correction algorithms. Given that important areas of the world have significant issues with decorrelation and atmospheric water vapor, interferometry research incorporated methods to reduce the problem to one more similar to continuous GPS. Corner reflectors have long been used as coherent calibration sources for interferometric systems, as they are observable over a wide range of angles, and have a well-defined phase center. Usai was the first to note that some man-made structures can behave much like corner reflectors, as though they are continuously reliable coherent scatterers over time and independent of the interferometric baseline (Usai, 1997). By identifying points in a series of radar images that maintain their coherence over time, we can create a network of phase measurements over time and space that are directly analogous to GPS measurements, though sampled once per month rather than continuously. Ferretti et al. (2000, 2001) were the first to systematize and popularize the treatment of these discrete point networks for long-term trend analysis. A number of groups have pursued similar techniques (Hooper et al., 2004, 2012; Wegmuller, 2003).

In practice, it is challenging to determine the location of time-coherent scatterers because the phase of each point is composed of a topographic component, displacement component, and noise components, all of which are different in each image of a series of images:

$$\phi\Delta t_i = \phi_{\text{topo},\Delta t_i} + \phi_{\text{def},\Delta t_i} + \phi_{\text{atm},\Delta t_i} + \phi_{\text{noise},\Delta t_i} \qquad [25]$$

where Δt_i is the time interval for interferogram, i, and the interferometric phase, $\phi\Delta t_i$, is broken into its principal components. Though the topography is assumed to be static over time, the phase term $\phi_{\text{topo},\Delta t_i}$ is variable because each interferogram has a different perpendicular component of the baseline $B_{\perp,i}$. Similarly, the other component terms will change over time. If we simply coalign all available images and then estimate the correlation over time for each point using eqn [20] (with expectation approximated by time averaging), then the variability of the phase over time will lead to an estimate of zero correlation. Ferretti et al. (2000, 2001) chose to identify time-coherent points by using a brightness threshold, as points with high mean brightness over time have small phase dispersion. This requires that images are radiometrically calibrated, such that variations in the brightness are due to scene variations, not radar system variations (e.g., radar antenna pointing variability).

Once these initial points are identified, it is necessary to find a solution for each phase term for all i that maximizes the correlation estimate. The maximization procedure involves a search over reasonable domains of expected phase values, which can be intensive. For example, if a DEM is available, the topographically induced phases are known, except for errors introduced by the DEM errors, so the search range for topography should be bounded in a region around the known topography by that error. The deformation is parameterized by a functional form such as linear or sinusoidal, and the solution of the parameters is part of the maximization. The atmospheric error is assumed to be spatially slowly varying and uncorrelated over time.

This technique has been employed with success in urban areas where man-made time-coherent scattering points abound. Results include the measurement of subsidence rates of individual buildings at the level of <1 mm year^{-1} and seasonal effects due to groundwater withdrawal and recharge (Colesanti et al., 2003; Ferretti et al., 2000). Applications to problems of geophysical interest are limited because the areas of interest are generally larger and sufficient time-coherent scatterers frequently do not exist in nonurban environments. Bürgmann et al. (2006) combined GPS-derived horizontal velocities and Permanent Scatterer InSAR estimates of uplift in the San Francisco Bay Area to track tectonic uplift in areas not subject to seasonal effects, at an accuracy of better than 1 mm year^{-1}.

Other methods have been developed to exploit long time series available for over a decade of SAR observations but without the restriction of using only time-coherent point scatterers and without specific assumptions of the model of deformation. By limiting data to only those scene combinations where the baseline is well below the critical baseline, one can expand the area of usable image pixels that are from natural surfaces but are coherent in time as well (Berardino et al., 2002; Hetland et al., 2012; Lanari et al., 2004b;

López-Quiroz et al., 2009; Lundgren et al., 2001; Schmidt and Burgmann, 2003). In these methods, each collection of images allows the generation of a dense spatial network of points (for well-correlated areas, potentially every image pixel could be included in the network). For each point, interferometric pairs formed from various combinations of images yield phase differences over available time intervals in the time series where correlation is good, with the shortest sampling interval being the satellite repeat period. From these phase difference measurements for each network point, an inversion is carried out to reconcile all differences. The inversion adjusts each interferogram by a constant to bring all differences into agreement and attempts to integrate the phase differences in a piecewise linear fashion, minimizing the distance between phase estimates from one time step to the next. The inversion can be carried out with constraints that smooth the estimates of phase. It is also possible to incorporate parameterized model functions in the inversion, for example, for seasonal effects, if desired (Hetland et al., 2012). These techniques derive directly from GPS network inversion methods, with the substantive difference being the much higher density of samples in space, but much lower density of samples in time, for InSAR. These approaches are further discussed later in the chapter.

Ferretti et al. (2011) had further developed point-scatterer methods to include distributed scatterers as well, identifying neighboring pixels with statistically homogenous properties as associated with a single phase to be estimated. Computing the temporal coherence across all possible interferograms allows consistent estimation of phase values for each distributed scatterer such that they can be treated in the processing chain as point scatterers.

3.12.3.3 Speckle Tracking and Pixel Tracking

InSAR provides the satellite LOS component of relative displacements within an image. Albeit at lower sensitivities, we can also measure the along-track component of the displacement field using cross correlation techniques. This approach is frequently referred to as speckle tracking, pixel tracking, or range offsets (ROs) and azimuth offsets (AZOs). For crustal deformation purposes, we frequently only consider the along-track or AZOs that are a purely horizontal component of the displacement field and are by construction perpendicular to the LOS phase-based observations. This approach is strictly speaking not an interferometric approach, but we include it here since it uses the same single-look complex (SLC) images used in forming interferograms (**Figure 15**).

At its most basic level, this approach is simply automatic tie-point estimation and frequently uses the same algorithms and software used to calculate tie-points when coregistering two SLCs to form an interferogram or to coregister a DEM-based height map to an interferogram to remove topographic effects (e.g., Michel et al., 1999a,b). Given a master and slave SLC and a rough guess at their relative gross offsets, a chip of $N \times M$ pixels (typically several tens of pixels on a side) is cross correlated from both images to provide range and AZO with subpixel accuracy. The complex or amplitude image is typically oversampled by a factor of 2 before the cross correlation to avoid spectral aliasing of the images. The peak of the cross correlation surface is identified therefore at half-pixel spacing

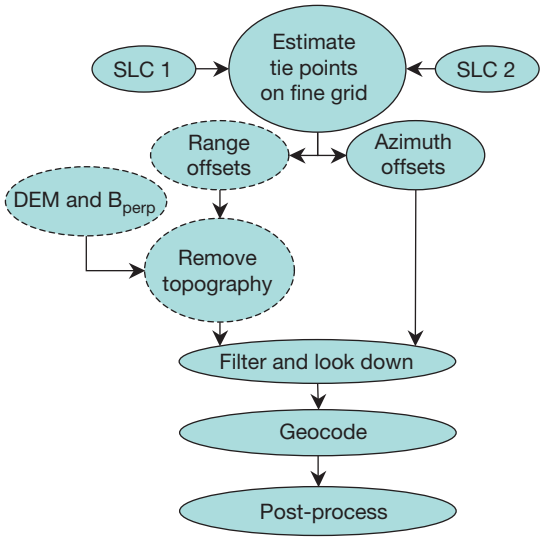


Figure 15 Generic processing flow diagram for generating pixel-tracking measurements. We assume here that the products of the associated differential InSAR processing have already been made.

as a first approximation to the offset location. The correlation surface is then interpolated to find the correlation peak with finer granularity. This process can then be repeated over a grid. Pixel tracking has deformation measurements that have the advantage that they do not need to be phase unwrapped. Since the measurement is done using ensembles of pixels, the spatial resolution is substantially degraded relative to standard interferometric observations. This sensitivity of the measurement is limited by the dimensions of a pixel. Typical sensitivities are about somewhere between one-tenth and one-thirtieth of a pixel dimension. For the ERS satellites, these numbers corresponded to about 10–20 cm, with better sensitivities for the high-resolution imagers such as TerraSAR-X, TanDEM-X, and the COSMO-SkyMed. While the range offsets provide the same component of the displacement field as the interferometric phase, it is usually much less sensitive and is therefore not usually used. However, there are times when correlation is low that the phase-based measurements will not work, but range offsets can (e.g., Peltzer et al., 1999). In such cases, the range offsets need to be corrected for topographic effects in a way similar to what is done for the phase measurements. The AZO measurement complements the interferometric phase measurements and therefore always provides useful observations when the expected displacements are large enough to be detected.

Algorithmically, the approach is the same as that used to generate tie-points or offset measurements using optical imagery (Michel et al., 1999a,b), although with InSAR data, we can achieve extremely accurate estimates of offsets with no inherent contrast in the mean radar backscatter. This is because SAR data exhibit high-contrast speckle characteristics from pixel-to-pixel (**Appendix A.1**) that are the same in interferometrically coherent SAR images, allowing good matching.

Cross correlation can be performed using the complex images directly or using just the amplitude of the complex image. When complex images are used, we are in essence computing small interferograms and assessing the quality of

the fringes therein. Because the phase is used, when the coherence is good, this method can lead to very tight constraints on the pixel offsets. However, when the coherence is poor, the phase contribution leads to extremely poor correlation estimates even when common surface features that are well correlated are present. This is truly a 'speckle-tracking' approach in that if the speckle differs between images, it will not track well. Cross correlation of amplitude images tracks features and so would be more accurately called 'feature tracking' or 'pixel tracking.' Because speckle appears as high-contrast features from pixel-to-pixel, amplitude matching also tracks speckle, though it lacks the tightness of the match of complex image correlation when the coherence is good. However, amplitude matching can perform well when images are rich with surface features in areas of poor interferometric correlation. Experimentation with both approaches shows that very little matching accuracy is lost by correlating amplitude rather than complex images.

We illustrate the use of pixel tracking with an example taken from the 1999 M_w 7.6 Chi-Chi, Taiwan, earthquake. This event was a thrust earthquake where the footwall had relatively little topography, while the hanging wall is extremely rugged. Standard InSAR produced clean fringes in the footwall region but is completely decorrelated in the hanging wall (Figure 16). In contrast, the AZO observations can be made for the entire image, with the caveat that strong spatial (median) filtering was required to remove outliers (Figure 17).

Instead of using speckle- or pixel-tracking methods as described in the preceding text to infer the along-track component of displacement, Bechor and Zebker (2006) and Jung et al. (2009) proposed splitting the aperture normally used in for a single interferogram into separate interferograms using the forward and backward squinting SLCs (relative to the nominal squint angle for the standard SLC). These two interferograms can then in turn be differenced to produce a map of along-track displacements. They showed accuracies on the order of a few centimeters when the interferometric coherence was very good. In areas of lower coherence, the phase difference-derived estimates were comparable to typical AZO estimates.

3.12.3.4 Other Advances

With the ever-increasing capabilities of spaceborne SAR systems, including finer resolution, polarimetric diversity, and tandem operations with multiple baselines, there is continued active development of new techniques and applications that have bearing on interferometry and geodesy. Systems such as TerraSAR-X and COSMO-SkyMed deliver fine-resolution imagery suitable for interferometric applications. With fine

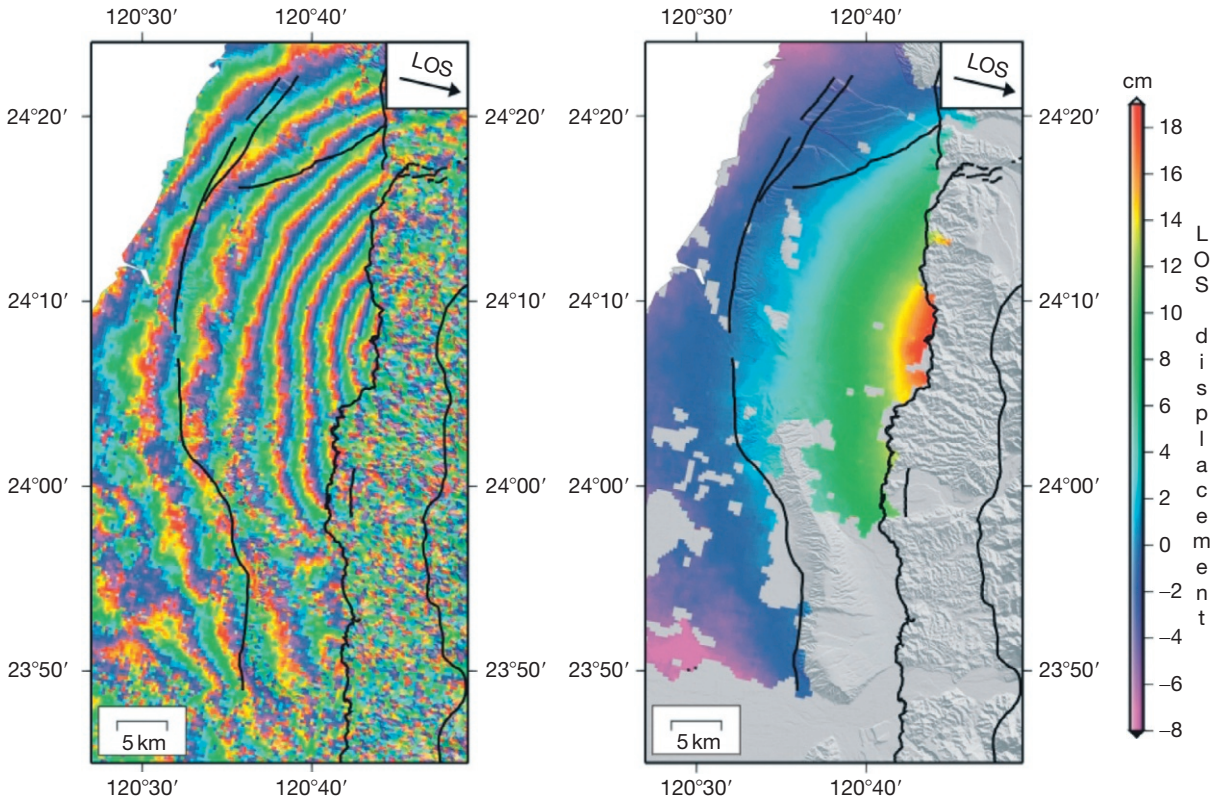


Figure 16 Wrapped at 2.8 cm per fringe (left) and unwrapped (right) interferogram for the 1999 M_w 7.6 Chi-Chi, Taiwan, continental thrust earthquake using ERS C-band data from 5 May 1999 to 23 September 1999. Black lines indicate major faults traces. The Chelungpu Fault is the primary slipping structure for the Chi-Chi, Taiwan, earthquake and forms the eastern limit of the unwrapped phase. The phase is successfully unwrapped only on the footwall where the topographic gradients and vegetation cover are dramatically less than in the hanging wall. Adapted from Levy F, Hsu Y, Simons M, LePrince S, and Avouac JP (2005) Distribution of coseismic slip for the 1999 Chi Chi Taiwan earthquake: New data and implications of varying 3D fault geometry. *Eos*, *Transactions of the AGU*, *Fall Meeting Supplement* 86: B1305.

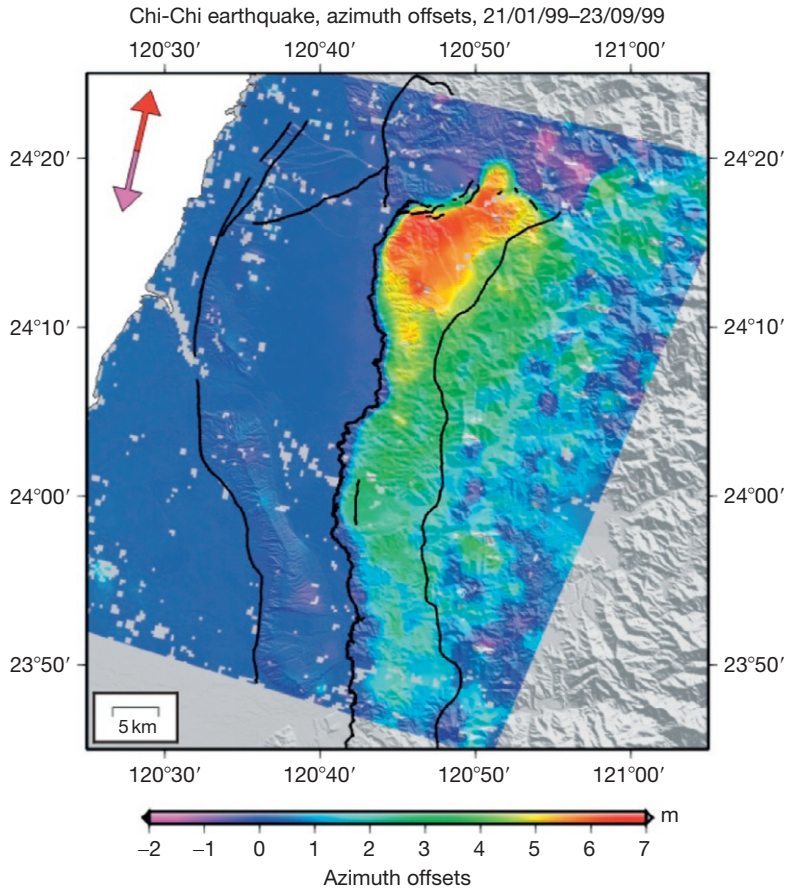


Figure 17 Azimuth offset-based horizontal displacements for the 1999 M_w 7.5 Chi-Chi, Taiwan, continental thrust earthquake derived from ERS *C*-band data from 21 January 1999 to 23 September 1999. Initial offset measurements are estimated using a correlation window of 23×160 pixels equivalent to a resolution of 320 m. A bilinear ramp component of the displacement field is constrained to match that estimated using available GPS data. These offsets are then spatially filtered with a 3 km wide median filter. Adapted from Levy F, Hsu Y, Simons M, LePrince S, and Avouac JP (2005) Distribution of coseismic slip for the 1999 Chi Chi Taiwan earthquake: New data and implications of varying 3D fault geometry. *Eos*, *Transactions of the AGU*, *Fall Meeting Supplement* 86: B1305.

resolution, single scatterers have a greater likelihood of dominating a pixel and therefore the density of coherently resolved points on the ground increases. This not only allows densification of the pattern of deformation in an area, but it allows one to individuate a small number of scatterers that dominate a pixel's response by applying tomographic techniques on a set of images acquired with a range of baselines (Zhu and Bamler, 2010). This could aid the geodesist in distinguishing ground motion from building motion in urban environments. In a similar fashion, in forested terrains, tomographic and multibaseline polarimetric techniques are being exploited to separate ground scattering from canopy scattering (Cloude, 2006; Cloude and Papathanassiou, 1998; Neumann et al., 2008). The tomography is used to distinguish vertical structure, with the polarimetry assisting in localizing different scattering mechanisms, which tend to be in different heights in the canopy (e.g., randomly arranged branches well above the ground versus regular ground-trunk structures at ground level). For the geodesist, successful isolation of the ground signature of deformation could potentially improve accuracy of time series. Though these tools are largely in the province of specialists, tools are becoming available for the community to explore (e.g., PolSARpro for polarimetric interferometry; http://earth.eo.esa.int/polsarpro).

3.12.4 A Best-Practices Guide to Using InSAR and Related Observations

Our general goal with geodetic imaging is to discover new crustal deformation processes and to estimate the value and uncertainties of controlling parameters associated with these as well as previously recognized mechanisms. The approaches adopted to using these data vary depending on the extent and type of observations available and the geophysical target. Here, we address some of the approaches that have found some utility in different situations. Beyond the exploratory mode of just looking at interferograms to discover things, we can consider how to best combine interferograms, deal with long time series, and how to incorporate these observations in parameter estimation schemes in a computationally efficient manner. In all cases, we must consider the different components of the measurement that act as apparent errors in the geodetic signal. From the perspective of modeling tectonic process, the primary

sources of error come from inadequacies in our knowledge of the satellite orbits and propagation delays accumulated in the ionosphere and troposphere. Here, we begin with a summary of error sources since that impacts all uses of the data. At a minimum, knowledge of the error structure (i.e., the covariance structure) is important in order to correctly execute any estimation of geophysical parameters. Furthermore, in certain circumstances, a given component of the error may be reduced using a variety of techniques.

3.12.4.1 Interferometric Baseline Errors

The primary impact of orbital errors is to induce long wavelength phase ramps associated with incorrect removal of topographic effects. If B_{perp} is long, then errors in B_{perp} can also lead to short wavelength errors in regions of rough topography. Typically, apparent long wavelength deformation gradients associated with errors in B_{perp} are dealt with by removing a best-fit bilinear or biquadratic polynomial ramp from both the observations and the models before comparing the two (e.g., Cavalié et al., 2008; Pritchard and Simons, 2006; Pritchard et al., 2006) or by using independent data, usually GPS observations, to either reestimate the interferometric baseline or to constrain the longest wavelengths of the physical model (e.g., Gourmelen et al., 2010; Peltzer et al., 2001; Simons et al., 2002). Alternatively, if the deformation signal is known to be localized within the interferogram, one can flatten the image by assuming zero deformation in the far field.

In order to merge GPS observations with InSAR data, we must take care how these data are combined. The surface deformation field frequently shows a strong quasi-periodic seasonal component (e.g., Dong et al., 2002). This seasonal signal is exemplified by the GPS time series in Figure 18, which shows a >1 cm peak-to-peak seasonal displacement. Of course, with InSAR data, one is sensitive to spatial gradients in these seasonal signals on the scale of the InSAR image, which will usually be smaller than implied by these kinds of GPS observations. If one uses GPS to refine estimates of B_{perp}, then one should use the GPS data before any seasonal signals have been removed. Conversely, if one assumes the long wavelength in the data is unconstrained and uses the GPS to constrain the model directly, then clearly, the seasonal signal should be removed first. This approach is safest since the seasonal signal in the GPS is not completely due to surface displacements, and similarly, there are long wavelength errors in interferograms caused by tropospheric delays and not by errors in B_{perp}.

3.12.4.2 Propagation Delays

The atmosphere and ionosphere introduce propagation phase and group delays to the SAR signal. In repeat-track systems, the propagation effects can be more severe. The refractive indices of the atmosphere and ionosphere are not homogeneous in space or time. For a spaceborne SAR, the path delays can be very large, depending on the frequency of the radar (e.g., >50 m ionospheric path delay at L-band), and can be quite substantial in the differenced phase that comprises the interferogram (many centimeter differential tropospheric delay, and meter-level ionospheric contributions at low frequencies). These effects in repeat-track interferograms were first identified by Massonnet et al. (1993) and later by others (Goldstein, 1995; Hanssen, 2001; Lohman and Simons, 2005; Massonnet and Feigl, 1995; Rosen et al., 1996; Tarayre and Massonnet, 1996; Zebker et al., 1997). Ionospheric delays are dispersive, so frequency-diverse measurements can potentially

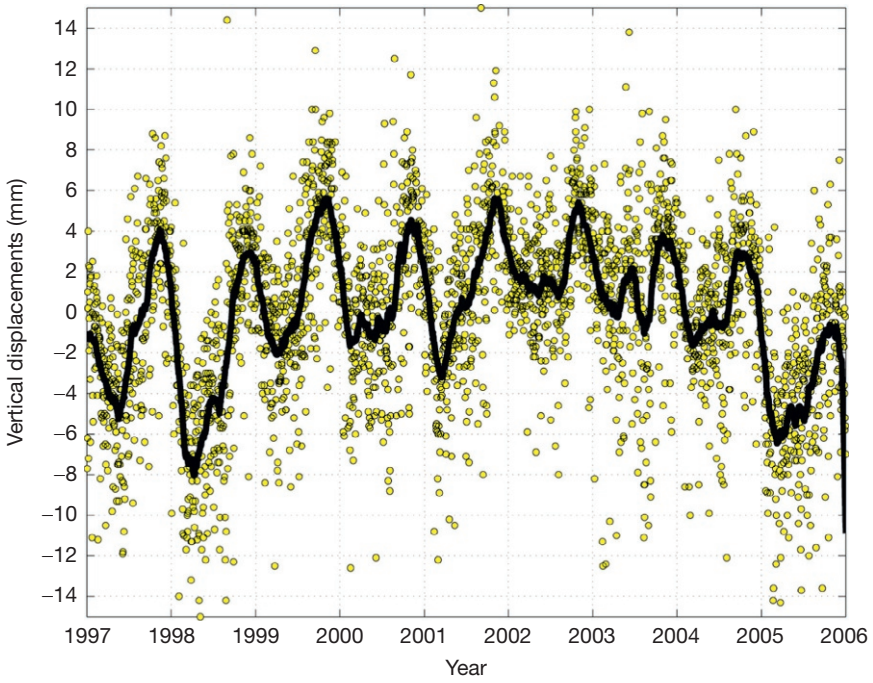


Figure 18 Dots indicate vertical GPS time series from the continuous GPS site TABL located on the central San Andreas Fault, CA. The displacements have been detrended and have the common mode removed. The solid line indicates the running 2-month average. Data from http://sopac.ucsd.edu.

help mitigate the effect, as with two-frequency GPS systems (Meyer, 2010). Tropospheric delays are nondispersive and can mimic topographic or surface displacement effects. Schemes for distinguishing tropospheric effects from other effects have been proposed (Massonnet and Feigl, 1995), and the use of averaging interferograms to reduce atmospheric noise is common (Fujiwara et al., 1998; Zebker et al., 1997).

Even temporal variations of a purely horizontally stratified troposphere will produce topographically correlated delays due to differences in total water content (Hanssen, 2001). Such variations are frequently loosely termed tropostatic effects as they do not include the impact of atmospheric turbulence. Given a simple model of the distribution of water and a few points of calibrations from GPS estimates of zenith wet delay (ZWD) or from radiosondes, one can attempt to correct for this effect (Delacourt et al., 1998). Sometimes, an empirical estimate of the linear relationship between phase and topography is warranted and can significantly improve the quality of an interferogram (e.g., Biggs et al., 2007; Cavalié et al., 2007, 2008; Lin et al., 2010; Remy et al., 2003; Shirzaei and Bürgmann, 2012). Given multiple interferograms with common scenes, it is important to make these estimates in an internally consistent manner, and in order to avoid bias from a single dominant scale of atmospheric anomaly or the impact of a extensive deformation signal, it may also be beneficial to infer the linear relationship through a robust (using L1 norms) and multiscale formalism (Lin et al., 2010). One may forgo the empirical linear or quadratic estimate of the relationship between elevation and phase delay and instead rely on predictions from regional or global atmospheric models (Doin et al., 2009; Elliott et al., 2008; Jolivet et al., 2011) such as ERA-Interim from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Dee et al., 2011) and North American Regional Reanalysis (NARR) (Mesinger et al., 2006).

Unfortunately, excursions from horizontal stratification are the norm, and in fact, in any given interferogram, it is common to see phase variations that correlate with topography but with amplitude (and even sign) varying from feature to feature. There are a variety of more involved approaches to dealing with the problem of path delays associated with spatial and temporal variations in tropospheric water vapor content. One can estimate statistically their impact on the estimates of tectonic deformation and account for these in the data covariances when modeling. Alternatively, one can attempt to explicitly model and remove these delays.

In terms of estimating the appropriate statistics, several studies have attempted to define the variance and distance-dependent covariance of these delays (e.g., Emardson et al., 2003; Goldstein, 1995; Hanssen, 2001; Li et al., 2004, 2006b; Lohman and Simons, 2005; Massonnet and Feigl, 1995; Williams et al., 1998; Zebker et al., 1997). Using estimates of ZWD from GPS data in Southern California, Emardson et al. (2003) assumed an isotropic (azimuth-independent) model and found that the variance, σ, between two locations varies as $\sigma = cL^{\alpha} + kH$, where L and H are the relative horizontal and vertical separation distances between locations, the values of c, α, and k are estimated to be 2.5, 05, and 4.8, respectively. The value of α is expected to be more or less region-independent, but the value of c may vary between regions. This model must then be scaled to account for the InSAR LOS direction. Limitations of this model include the assumption that the height dependence is independent of absolute height and that there are no atmospheric variations at scales smaller than the spacing in the GPS networks, which could influence the estimate of the height dependence (Xu et al., 2006). Lohman and Simons (2005) found reasonable agreement between the estimates from Emardson et al. (2003) with empirical estimates from real interferograms (**Figures 19** and **20**). Adopting this statistical approach assumes that one has a large ensemble of interferograms to use in ones analysis. If one has only a few image pairs, then the statistics of small numbers comes into play when interpreting a specific feature in a given interferogram. Furthermore, if one removes the long wavelengths from of a given interferogram, then one is also affecting the estimates of σ. These difficulties suggest that at a minimum, one should estimate the full covariance empirically, if necessary, removing an initial model of the phenomena of interest first (Lohman and Simons, 2005).

Since GPS can measure both the ZWD and its spatial gradient, it is possible to use the GPS observations to make a wet delay image to correct individual interferograms (Li et al., 2004, 2006b; Lofgren et al., 2010; Onn and Zebker, 2006; Williams et al., 1998; Xu et al., 2006). This approach requires estimates of the delay made at or near the time of acquisition of the radar images. The viability of this approach clearly increases with density of GPS sites. In order to increase the spatial resolution of the images of wet delay beyond just the station spacing, Onn and Zebker (2006) demonstrated the use of time series of ZWD combined with a frozen flow assumption and observations of wind speed. Essentially, by assuming that the pattern of the wet delay is changing slowly, they can use estimates of ZWD before and after the image acquisition times to predict variations between GPS stations. To date, these approaches have not taken advantage of the spatial gradient of the ZWD that can also be estimated when processing the GPS delay. The spacing of the GPS network will always remain a fundamental limitation; however, since the spatial distribution of the wet delay decays with decreasing wavelength, any correction at medium and long wavelengths should significantly reduce effects due to tropospheric delays. Unfortunately, there are only a few, albeit important, regions of the world where the GPS network is sufficiently dense to make any of these methods viable.

The lack of dense GPS networks in all regions of geodetic interests motivates the need for satellite-based estimates of the wet delay. Significant progress has been made to this end using observations from near-infrared imaging radiometers such as NASA's Moderate Resolution Imaging Spectroradiometer (Li et al., 2005) and the MERIS instrument on ESA's Envisat satellite (Li et al., 2006a, 2012; Puyssegur et al., 2007). The latter has the advantage of estimating the water delay in the same viewing geometry and at the same time as the acquisition of the radar data. Unfortunately, imaging radiometers are limited to use in the daytime and cloud-free conditions. However, if clouds are not pervasive, then the resulting holes can be interpolated. Thermal infrared measurements could potentially be used for nighttime observations. Ideally, the use of image-based estimates would be done in combination with GPS estimates of wet delay where and when available.

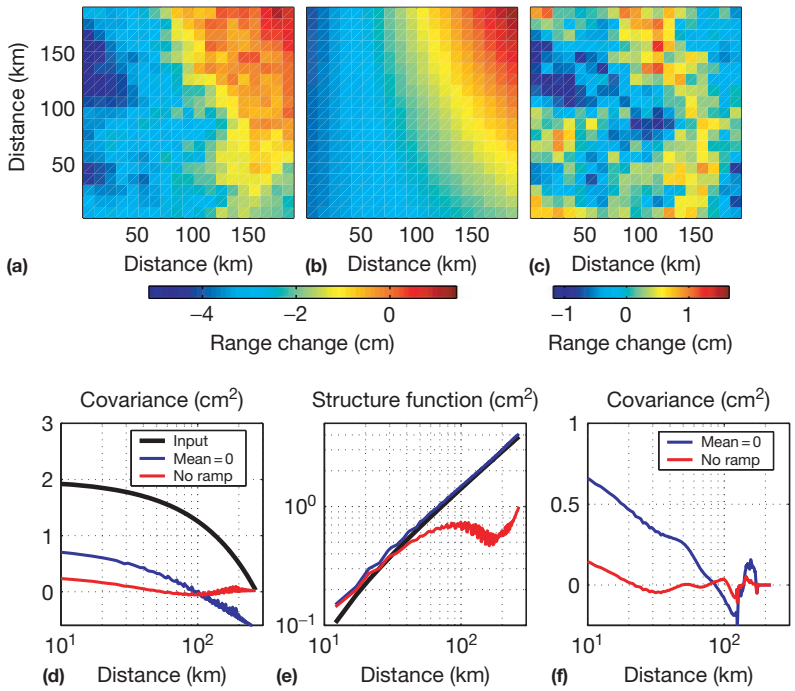


Figure 19 Effect of ramp removal on character of noise. (a) One realization of noise with power law structure function. (b) Best-fit quadratic ramp to (a). (c) Remaining signal after ramp removal. (d) Input covariance (thick line) and inferred covariance for synthetic noise in (a) after removal of mean (blue) and quadratic ramp (red). (e) Structure function for same scenario. (f) Covariance inferred for a real interferogram, after removal of the mean and with and without the removal of the best-fitting quadratic ramp. Reproduced from Lohman RB and Simons M (2005) Some thoughts on the use of InSAR data to constrain models of surface deformation. *Geochemistry*, *Geophysics*, *Geosystems* 6. http://dx.doi.org/10.1029/2004GC00084.

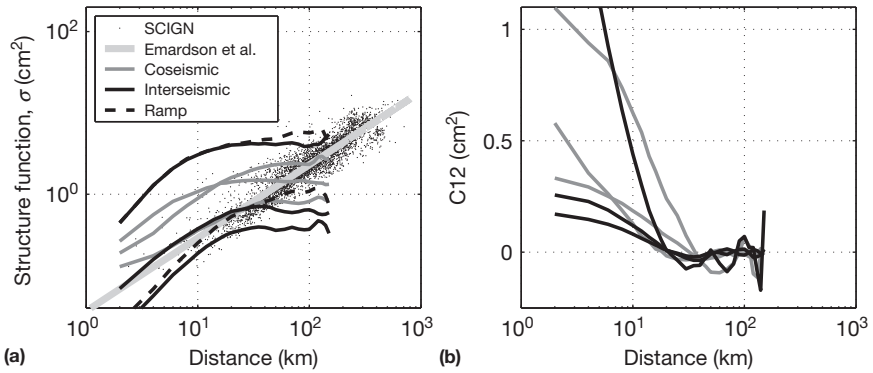


Figure 20 Spatial structure function versus distance for interferograms spanning the Hector Mine, California, earthquake (black) and interferograms spanning time periods before or after the earthquake (gray). Dashed lines indicate the structure function for the interferogram before a quadratic ramp was removed. GPS data (small dots) are from the Southern California Integrated GPS Network (SCIGN). The thick gray line indicates the power law fit. Adapted from Emardson TR, Simons M, and Webb FH (2003) Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB001781. (b) Covariance versus distance for the coseismic and interseismic interferograms from (a).

Thus far, we have only addressed direct measurements of the wet delay. Recently, progress has been made using high-resolution weather models to estimate the spatial variation in wet delay (Foster et al., 2006; Puyssegur et al., 2007). Foster et al. (2006) demonstrated the use of the MM5 weather model (Grell et al., 1995) to correct interferograms from the Big Island of Hawaii. In this case, the model is updated every 12 h using local meteorological data, and estimates of wet delay are made every hour and interpolated to the time of acquisition for each radar image. Using this model, they can generally reduce the variance by about a half at wavelengths of 30 km, although they find limitations in the model in regions of high topographic gradient. Unfortunately, a similar study in the region of Mount St. Helens (Washington) was less successful. Puyssegur et al. (2007) combined the MM5 model with MERIS observations. In the near future, where available, imaging radiometer and

GPS wet delay observations will be combined with physical models to produce the best estimates of wet delay possible. Indeed, one would expect that any future dedicated InSAR mission would produce and distribute estimates of wet delay for each image acquired. Clearly, once the best estimates of wet delay are removed, it is straightforward to empirically estimate the residual covariance structure on an interferogram-by-interferogram basis (Lohman and Simons, 2005).

3.12.4.3 A Simple Data Covariance Model for Multiple Interferograms

The full covariance can be constructed of three primary parts: (1) the variance of a given pixel on the diagonal, (2) the intraimage pixel-to-pixel covariance primarily due to ionospheric and tropospheric path delays, and (3) intraimage covariances associated with the use of a single common image in two separate interferometric pairs (Emardson et al., 2003). Assuming ergodicity, one estimates the pixel variance empirically using estimates of the local phase variance over a small patch of pixels. Doing this for each image is important since different interferograms or offset images may have had different amounts of spatial filtering applied to them as well as different amount of temporal decorrelation. The intraimage pixel-to-pixel covariance should also be computed empirically since the degree of filtering and long wavelength ramp removal may vary from image to image. Usually, this covariance is assumed to depend only on the relative distance between two given pixels (Emardson et al., 2003; Hanssen, 2001; Lohman and Simons, 2005). In contrast to the intraimage pixel covariance, the inter-image covariance can be derived analytically (Emardson et al., 2003).

3.12.4.4 Impact of Solid Earth and Ocean Tides

Measurements of ground deformation, be it from ground-based or space-based geodesy, will contain effects associated with solid Earth tides and the response of the solid Earth to ocean tidal loads (OTLs) (for a review, see **Chapter 3.06**). In conventional GPS analysis, these effects are normally modeled out. For the OTL effects, an oceanographic model of the time-varying OTL is assumed and its impacts on a spherically symmetric elastic Earth are calculated and removed from the observations to isolate the nontidal component of the deformation. Furthermore, in many GPS analyses, positions are averaged over a 24 h period thereby further mitigating the impact of the primary tidal constituents (diurnal and semi-diurnal) as well as any error on the tidal models.

By their very nature, InSAR observations are effectively snapshots and cannot be averaged over tidal periods. Furthermore, even though space-based SARs typically repeat their observations at the same local time, the tides do not repeat at exactly this period. Thus, any arbitrary collection of SAR scenes will sample all phases of the tides, and we must consider their impact on our observations. The effects of solid Earth tides are sufficiently well understood that they can be easily accounted for or even neglected given their very long wavelength nature. However, the elastic response to OTL should not be neglected, especially in regions within a few hundred kilometers of coastlines experiencing large tides (Dicaprio and Simons, 2008). For instance, there are >1 cm per 100 km gradients in the OTL induced vertical deformation associated with the M2 tidal constituent for regions near the coast of Cascadia, a scale and amplitude of deformation similar to that expected from interseismic deformation along any given subduction zone. A factor of 2 larger OTL response is expected near coastal Peru and across Iceland (Dicaprio and Simons, 2008).

3.12.4.5 Stacking Single Component Images

Independent of whether or not one can correct interferograms for path delay effects, if the primary geophysical target is a single event that occurred quickly (in terms of satellite InSAR, this implies anything lasting less than the orbit repeat time) or alternatively is a gradual process occurring at constant rate, it may be useful to increase the SNR ratio by stacking (i.e., averaging) multiple interferograms. Such stacks have the advantage of reducing effects due to tropospheric delays since these are uncorrelated on timescales of more than a day (e.g., Emardson et al., 2003; Hanssen, 2001) and thus is useful to discover small signals (e.g., Cavalié et al., 2008; Gourmelen and Amelung, 2005; Lyons and Sandwell, 2003; Peltzer et al., 2001; Pritchard et al., 2006). Stacking images can also reduce computational burden in parameter estimation schemes by reducing the amount of observations (e.g., Pritchard et al., 2006).

These stacks can be made either in radar coordinates (choosing one interferogram as the master) or in geocoded (latitude and longitude) coordinates. Before stacking, one needs to account for long wavelength errors or phase ramps associated with errors in estimates of the interferometric baseline for each interferogram. The phase ramp is usually parameterized as a bilinear or biquadratic polynomial function of azimuth and range (or geographic coordinates). Neglecting to deal with the ramp can result in biased stacks. If several GPS displacement vectors are available for the region of interest, then they can be used to resolve the ramp parameters either before or after forming the stack.

The simplest and most common brute force stack is made by just adding all the interferograms together. In this approach, regions of decorrelation in the resultant stack will be the union of decorrelated regions in the individual interferograms. Therefore, one may choose not to include an individual interferogram if it has an excessively large amount of decorrelation. An example of such a single LOS image stacks for a short-duration event comes from images of the coseismic displacements from deeper and smaller earthquakes in Chile (Pritchard et al., 2006). In most cases, it is advisable to take a more formal approach to stacking.

When considering making stacks, one should separate the case of a single rapid event from that of linear secular deformation. In the case of a single rapid event, the stack is made in terms of displacements. Scenes are averaged together in a straightforward fashion – ideally, using a weighted average, where the weighting is the inverse covariance matrix of the observations. When calculating a displacement stack with the full covariance matrix, one would normally pose the problem as a least squares problem. The use of a purely least squares approach relying on an L2 norm is somewhat debatable since phase noise is not necessarily Gaussian. It may in fact be more robust to use an L1 norm which is equivalent to asking for the

median and not the mean displacement. Regardless of the adopted norm, using a formal approach that includes the local estimate of variance permits one to include all available data regardless of the level of decorrelation.

For a constant rate process, the estimation of the rate at any given pixel is best described using a bit more formality. The underlying equation is simply

$$\rho_i = T_i v + \varepsilon_i \quad [26]$$

where ρ_i, T_i, v, and i are the observed range change, time span, rate, and error, respectively, for a pixel in the ith image. We can rewrite this equation as the linear system for the ensemble of images:

$$d = \theta v + E \quad [27]$$

where d, θ, and E are the vectors of observations, time spans, and errors, respectively. This system has a general weighted least squares solution

$$v = (\theta^t W \theta)^{-1} \theta^t W d \quad [28]$$

where, as described earlier, the weight matrix, W, is the inverse of the full data covariance matrix. This approach naturally deals with regions that are decorrelated in some but not all of the images. The same issue described above relating to L2 versus L1 norms applies here.

While not recommended, one could assume that the pixel variances are constant within a given displacement image and between different displacement images as well as ignore all the covariances. These frequently adopted but unnecessary simplifications lead to the explicit least squares solution:

$$v = \frac{\sum_{i=1}^{N} \Delta T_i \Delta \rho_i}{\sum_{i=1}^{N} \Delta T_i \Delta T_i} \quad [29]$$

In a study of postseismic deformation from historic earthquakes in the Basin and Range Province of the Western United States, Gourmelen and Amelung (2005) removed the long wavelength ramp from individual component interferograms, then add the residual interferograms up, and divide the total displacement by the summed total time span of all the interferograms. This approach is equivalent to converting all the individual interferograms to rates and then averaging the individual rates, weighted by their respective time spans. In other words,

$$v = \frac{\sum_{i=1}^{N} \Delta \rho_i}{\sum_{i=1}^{N} \Delta T_i} = \frac{\sum_{i=1}^{N} \left(\frac{\Delta \rho i}{\Delta T_i} \right) \Delta T_i}{\sum_{i=1}^{N} \Delta T_i} \quad [30]$$

Equation [30] is only the same as the least squares solution shown in eqn [29] if all the interferograms span equal time periods. In that special case, then

$$v = \frac{1}{N \Delta T} \sum_{i=1}^{N} \Delta \rho_i \quad [31]$$

In a study of slow interseismic deformation in the eastern California shear zone, Peltzer et al. (2001) effectively adopted eqn [31]. Given that they use interferograms of approximately the same duration (about 4 years), they simply averaged the rates (unweighted by time), which is close to the least squares solution assuming constant and uncorrelated data errors.

In general, given today's computational resources, there is no reason to not use the full weighted least squares estimation (e.g., eqn [28]). In particular, this approach allows a more rigorous estimate of both rates and their associated errors. When a final stack is complete, if desirable, it is then possible to create a mask based on the final estimated variances. More generally, the stacking process for either a point event or a constant velocity can and should be viewed as a particular simplified instance of more general time-series approaches.

3.12.4.6 InSAR Time Series

We can adopt more complicated models than the constant rate or single step function just described. For instance, given the potential for seasonal signals (e.g., **Figures 1** and **18**), it may be desirable to explicitly include a seasonal variation in the estimation process. Indeed, given enough temporal sampling, there exists the possibility for a wide range of time-series approaches effectively mimicking all the approaches adopted in GPS analysis. One approach aims to make the cleanest time series possible for later mechanical modeling. This approach can include estimates of site-specific signals including seasonal signals (frequently assumed to be sinusoidal), random noise at each site, and spatially correlated processes such as daily variations in reference frame estimates (e.g., Dong et al., 1998, 2003, 2006). Generally, the temporal and spatial analyses are separated in these approaches. A more complex underlying temporal evolution is also sometimes adopted in GPS analysis including a superposition of behaviors including linear rates, coseismic steps, and postseismic exponential or logarithmic decays. At a minimum, these more complicated model terms reduce biases in estimating seasonal and reference frame contributions and can always be added back in for subsequent mechanical modeling. Of course, the inferred model terms, such as coseismic steps, are themselves useful as input into subsequent mechanical models. An alternative approach simultaneously solves for time-varying parameters of a specified mechanical model and the aforementioned nontectonic signals (e.g., McGuire and Segall, 2003; Segall and Matthews, 1997).

The parallels between GPS and InSAR time-series analysis are numerous. Obviously, the physical processes of interest are the same, and they are sensitive to similar nontectonic processes such as seasonal effects. In addition, the ambiguity in absolute displacements with InSAR data is equivalent to reference frame error in GPS. However, InSAR time-series analysis has a few unique challenges. In terms of nontectonic signals, we may want to remove the effects of the troposphere (empirically or by physical model) and we have errors due to inaccurate removal of topographic effects stemming from inaccurate orbits and errors in the topographic data. With typical suites of interferograms from a given orbital track, we may also face incompletely connected chains of dates. For instance, given images acquired at times A, B, C, and D, we may be able to form interferograms I_{AC}, I_{BC}, and I_{CD}, or perhaps, we only have

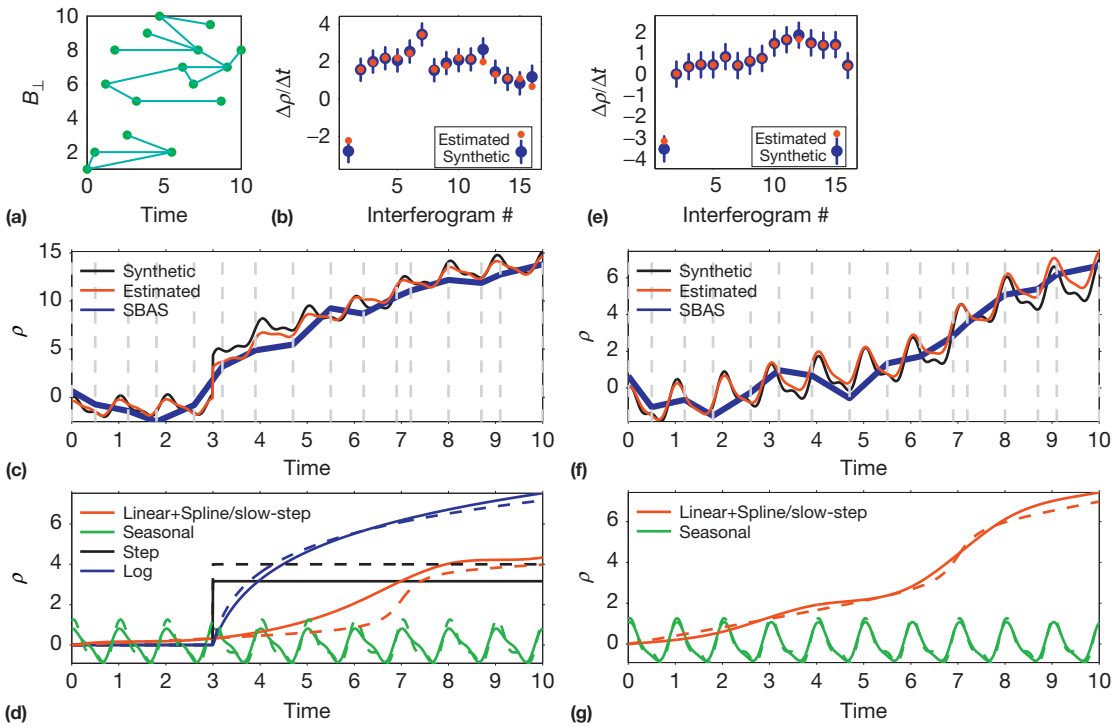


Figure 21 Illustration of different time-series approaches. (a) Schematic nondimensional $B_\perp$ versus time plot to illustrate the typical scenario for constructing interferograms from a given image archive. This figure also illustrates the idea of an 'island' of unconnected pairs whose connection to the other pairs is ambiguous. The pairs shown here are used to generate the results shown in the other panels. (b) Effective rate for each difference measurement derived from the synthetic in (c) after adding noise. (c) Input synthetic time series (black line), reconstructed time series using small baseline subsets (blue line) and physically parameterized approach (red line). Times of image acquisition are indicated by dashed gray lines. (d) Input (dashed) and reconstructed (solid) component signals. (e)–(g) Same as (b)–(d) for a simpler input model. Figure courtesy of Eric Hetland.

I_{AB} and I_{BC}. In the later case, we would like to be able to construct a continuous time record (**Figure 21**).

To deal with these issues, we have the same variants in modeling approaches for InSAR time series as just described with GPS data. One approach attempts to decompose a suite of interferograms into its component time intervals (e.g., AB, BC, and CD) and to solve for the incremental displacements. Ambiguities in combining unconnected pairs (or collection of pairs) require some form of regularization of the time series. One approach, called the small baseline subsets based on use of singular value decomposition (SVD), finds a time series that fits the available interferograms while minimizing the implied velocities in any given underlying time interval (Berardino et al., 2002). This approach is equivalent to minimizing the temporal gradient of the deformation field (**Figure 21**). Alternatively, one could use explicit Laplacian damping to minimize the roughness of the temporal evolution. The approach described by Berardino et al. (2002) begins with unwrapped interferograms and assumes that they have been tied to some stable reference point to define a common phase bias. This approach is a pixel-by-pixel algorithm at its core (important for computational purposes), but they apply subsequent spatial and temporal filtering to provide cleaner time series. An SVD or principal component approach may also be useful to isolate atmospheric signals in large InSAR time series from spatially and temporally coherent deformation fields (Ballatore, 2006). Examples of models using large time series of InSAR data include deformation at volcanoes (Lundgren et al., 2001, 2004), in the Santa Clara Basin (California) (Schmidt and Burgmann, 2003), in the Los Angeles Basin (California) (Lanari et al., 2004a), faulting in the Asal Rift, Djibouti (Doubre and Peltzer, 2007), or urban subsidence in Mexico City (López-Quiroz et al., 2009).

There are several other variants of InSAR time-series analyses (e.g., Biggs et al., 2007; Hetland et al., 2012; Shirzaei, 2013; Usai, 2003). For instance, in a method called Multiscale Interferometric Time Series (MInTS), Hetland et al. (2012) combined a temporal parameterization that is physically based (e.g., seasonal, secular, steps, and functional forms for transients associated with earthquakes) with the possibility to augment the temporal parameterization to include functions (e.g., splines) that allow one to extract signals from transients without a specifically known temporal behavior (**Figure 21**). Obviously, for coseismic steps, it is best if there are data closely bracketing the event. There will always be a trade-off between various components if there are no extensive periods with data before 'unusual' events. Given ensembles of interferograms with sufficiently dense temporal sampling, the seasonal signal is usually well estimated and therefore readily removed, as are any ambiguities with unconnected islands of pairs. The complexity of this approach stems from the choice of which suite of functions to adopt and how to regularize or damp their behavior in an appropriate way.

While most time-series analysis approaches consider each pixel separately, MInTS also proposes a wavelet basis to

parameterize the deformation field in space. Thus, the time-series analysis is actually done on spatial wavelet coefficients and not pixels. The use of a wavelet basis allows one to consider ensembles of interferograms with varying amounts of holes due to decorrelation. Holes are interpolated before transforming into the wavelet domain, and the relative amount of holes to data in estimating a given wavelet coefficient is used as a weight in the time-series analysis. Thus, in a given interferogram, a wavelet that is constructed mostly of interpolated data will contribute very little to the final analysis. In this manner, small holes are effectively interpolated using neighboring information in both time and space. Typically, the pixel-by-pixel approach ignores the established spatial covariance between pixels (e.g., **Figure 20**) since that would make the estimation problem computationally very difficult. In contrast, parameterizing the problem with wavelets effectively diagonalized this covariance matrix, thus allowing one to more faithfully incorporate a realistic error model while doing each wavelet independently (Hetland et al., 2012).

There is now an open-source framework allowing for the comparison of different approaches to correcting for tropostatic effects, for the comparison of different time-series estimation methods, and for easy introduction of new methods all within a uniform structure. Called the Generic InSAR Analysis Toolbox (GIAnT), these tools are freely available online at http://www.earthdef.caltech.edu (Agram et al., 2013).

As with GPS data, an alternative approach to time-series modeling involves explicit use of a physical model. For example, Pritchard and Simons (2006) studied time-dependent postseismic slip and aseismic slip on the subduction zone megathrust in northern Chile assuming a simple model of time-dependent slip on a fault plane and imposing Laplacian smoothing in time in order to tie together unconnected groups of interferograms. A similar approach was adopted by Grandin et al. (2010) in a study of time-dependent rift opening in the Afar (Ethiopia). Alternatively, one could use a Kalman filter approach (e.g., Shirzaei and Walter, 2010). Regardless, the mechanical modeling approach benefits directly from the availability of any additional temporally continuous data such as can be provided by GPS observations.

3.12.4.7 Vector Displacements

Another form of stacking can be useful when data from multiple viewing geometries are available. In particular, we may want to (1) compare geodetic imaging data directly with other single component geodetic data, such as leveling or electronic distance measurements, without going through a physical model; (2) get a better grasp on what are purely vertical versus horizontal displacements; (3) as in the single component stacks described in the preceding text, combine as many images of the geophysical event to reduce noise and look for unexpected (undiscovered) processes; and (4) reduce the total amount of data used in a parameter estimation task. In these cases, it may be useful to combine displacement images from three or more viewing geometries to construct the best estimate of the east, north, and vertical displacement field. The input data need not be homogeneous in type. For instance, one could combine right-looking images from ascending and descending orbits and pixel offset estimates from one or more tracks to resolve the full 3D displacement field (Fialko et al., 2001a, 2005). Similarly, one could use observations from different overlapping beams of adjacent orbital tracks (Wright et al., 2004). The whole system would then be weighted by the covariance W described in the previous section. If sufficient data are available, then one can explicitly include estimation of ramp parameters, which, as before, are best constrained if at least a few independent geodetic observations are available. Estimates of the full 3D coseismic displacement field exist for several earthquakes including the 1999 M_w 7.1 Hector Mine, California, earthquake (Fialko et al., 2001a) and the 2003 M_w Bam, Iran, earthquake (Fialko et al., 2005) (**Figure 22**). With respect to future mission design, Wright et al. (2004) pointed out the reduced sensitivity of 3D deformation maps to north–south motions for missions with near-polar orbits.

While we may not need to explicitly make 3D decompositions of the displacement fields, having multiple components of the displacement fields is clearly important to constrain physical models such as earthquake and volcano source models (Fialko et al., 2001b; Lohman et al., 2002). For example, for all earthquakes models, when only one LOS component is observed, there will be a trade-off between the amplitude and rake of the fault slip. This trade-off will be further exasperated for small earthquakes if one does not know the location well. Generally, we note that it may be dangerous to oversimplify the problem by assuming a priori that the deformation field is either purely vertical or purely horizontal in nature. For instance, in large strike-slip systems, there is a temptation to assume that the near-fault displacement field is purely fault-parallel. However, even the largest strike-slip systems frequently have vertical or fault normal deformation associated with them that is either tectonic in origin (e.g., Klinger et al., 2005) or due to hydrologic effects associated with faults acting as flow barriers (Bell et al., 2002). The presence of any vertical displacements is particularly problematic, given the high sensitivity to vertical deformation in most InSAR data. Indeed, any ignored vertical displacement will corrupt estimates of the horizontal displacement, where the amplitude of this corruption is amplified by at least $1/\tan\theta$, where θ is the angle of incidence. This effect increases as the fault strike approaches the azimuth of the orbital track.

3.12.4.8 The Luxury of Sampling: Rationale and Approach

Geodetic imaging can have a very large number of observations – of order 10^6 pixels. In many cases, one wants to use these observations to constrain parameters from a mechanical model, such as distributed fault slip in an elastic half-space. In practice, the computational expense of the inverse problem can be computationally bound by the number of Green's functions one needs to calculate. In the case of linear problems where the source geometry is known, the Green's functions may not be known analytically (e.g., as with 3D variations in material properties) or one may still want to impose nonlinear constraints; both of these requirements can be expensive to implement. Even worse, for nonlinear problems, such as when we do not know the geometry of the source, we must recalculate the Green's functions at each iteration.

Having spatially continuous observations provides us the opportunity to use a selected subset of these observations. This

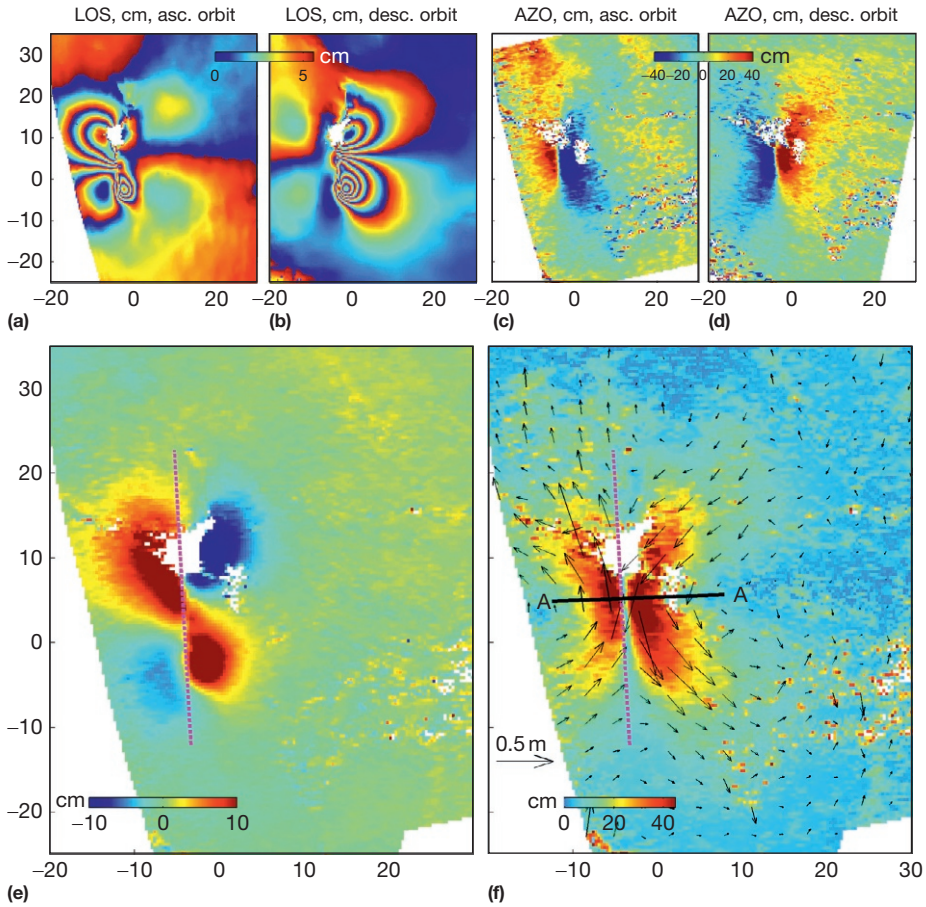


Figure 22 Phase and offset measurements for the 2003 M_w 6.5 Bam, Iran, shallow strike-slip earthquake. Images are from the C-band Envisat ASAR instrument. The coordinate axes are in kilometers, with the origin at 58.4 E, 29 N. Colors denote displacements in centimeters. (a) Interferogram for 16 November 2003 to 25 January 2004, ascending orbit. (b) Interferogram for 3 December 2003 to 11 February 2004, descending orbit. (c) Azimuthal offsets, ascending orbit. (d) Azimuthal offsets, descending orbit. Derived vertical (e) and horizontal (f) components of the surface displacement field. Arrows show the subsampled horizontal displacements. Dashed line shows the surface projection of the fault plane inferred from the inverse modeling of the data. Reproduced from Fialko Y, Sandwell D, Simons M, and Rosen P (2005) Three-dimensional deformation caused by the Bam, Iran, earthquake and the origin of shallow slip deficit. *Nature* 435: 295–299.

possibility leads to the question of the optimal subset of observations to pick. It is most precise to view this problem as spatially variable averaging and sampling. Proposed approaches fall in two classes, in the first, the image sampling is done based on properties of the data itself (Jonsson et al., 2002; Simons et al., 2002), and in the second, the sampling is controlled by the character of the model (Lohman and Simons, 2005). Jonsson et al. (2002) and Simons et al. (2002) proposed similar approaches that rely on successive subdivision of the deformation image. In both approaches, the image is cut into quadrants and a low-order best-fit surface is removed from the phase field in each quadrant. For a given quadrant, if the residual is greater than a prescribed threshold, the quadrant is further divided into four new quadrants, with the process allowed to continue until a minimum quadrant size is reached. For both approaches, one should then use the mean (or median) of the pixels in the quadrant and assign the value to the center of the quadrant. Jonsson et al. (2002) and Simons et al. (2002) differ in that the former considered the residual after removing the mean from each quadrant, while the latter removed a bilinear function. Both approaches work well, but for the same number of points, the Simons et al. (2002) approach does a better job at constraining a given fault slip model. Underlying the Simons et al. (2002) approach is the recognition that any physical model can at a minimum produce a linear deformation ramp and thus the ability to constrain detailed behavior of a given model lies in the curvature of the deformation field. In essence, we are attempting to choose a data set that includes as many points as possible while maintaining a nearly diagonal data resolution matrix. Both of these approaches are limited by the inherent noise in the observations and can give spurious regions of high sampling (e.g., far away from a target fault) if there are unwrapping errors or regions of decorrelation.

An alternative approach to image sampling involves using a best-guess initial model parameterization (Lohman and Simons, 2005). A given linear model can be written as $Gm = d$, where G is the design matrix, m is the vector of model parameters, and d is the vector of observations. This has a generalized solution $m_{est} = G^{-g}\mathrm{d}$, where G^{-g} is the

generalized inverse and the model and data resolution matrices, R and N, can be written as $R = G^{-g}G$ and $N = GG^{-g}$ (e.g., Menke, 1989). Lohman and Simons (2005) proposed to find a distribution of samples that has the most points while keeping N nearly diagonal. The algorithm is similar to that of Simons et al. (2002) and Jonsson et al. (2002) in the use of successive division into quadrants, with the difference being that at each stage, an estimate of N is made, and refinement stops when N becomes sufficiently nondiagonal. As with the other approaches, for any given quadrant, this approach uses the weighted mean (or median) value, where the weight takes into account the pixel variances and covariances. This approach is ideal if one has a reasonable first guess at the model geometry. It is of course also sensitive to the assumed model parameterization – in this case the size of each fault patch. Regardless, all these schemes permit one to constrain models with about 1% of the original data without significant loss of information. It is worth emphasizing that even where the final sample spacing is large, the data variance will be relatively low, since more points go into this estimate than in regions of finer spacing, thus information is not really lost.

These spatially variable averaging/sampling approaches are particularly important when modeling shallow sources with the potential of causing complex deformation patterns. A common example of this class of problems comes from shallow earthquakes. **Figure 23** demonstrates the difference in sampling patterns that result from the different approaches described here. This figure also demonstrates the difference in the ensuing model resolution. The choice of algorithm clearly impacts the ability to distinguish variations in the shallowest parts of the model (Lohman and Simons, 2005).

3.12.4.9 Decorrelation as Signal

Interferometric temporal decorrelation is usually viewed at least as a nuisance and sometimes as a complete barrier to making useful displacement measurements. However, in some cases, the spatial distribution of decorrelation, as well as its temporal evolution, can serve as useful indicators of geologic processes. There are many causes of temporal decorrelation that are not necessarily of interest; among these are the impact of weather-related processes (rain, snow, etc.). Of more interest is the decorrelation caused by rearrangement of the scatterers within a pixel associated with intense shaking, damage, etc. Examples of the use of such decorrelation include mapping of active lava flows (Zebker et al., 1996)

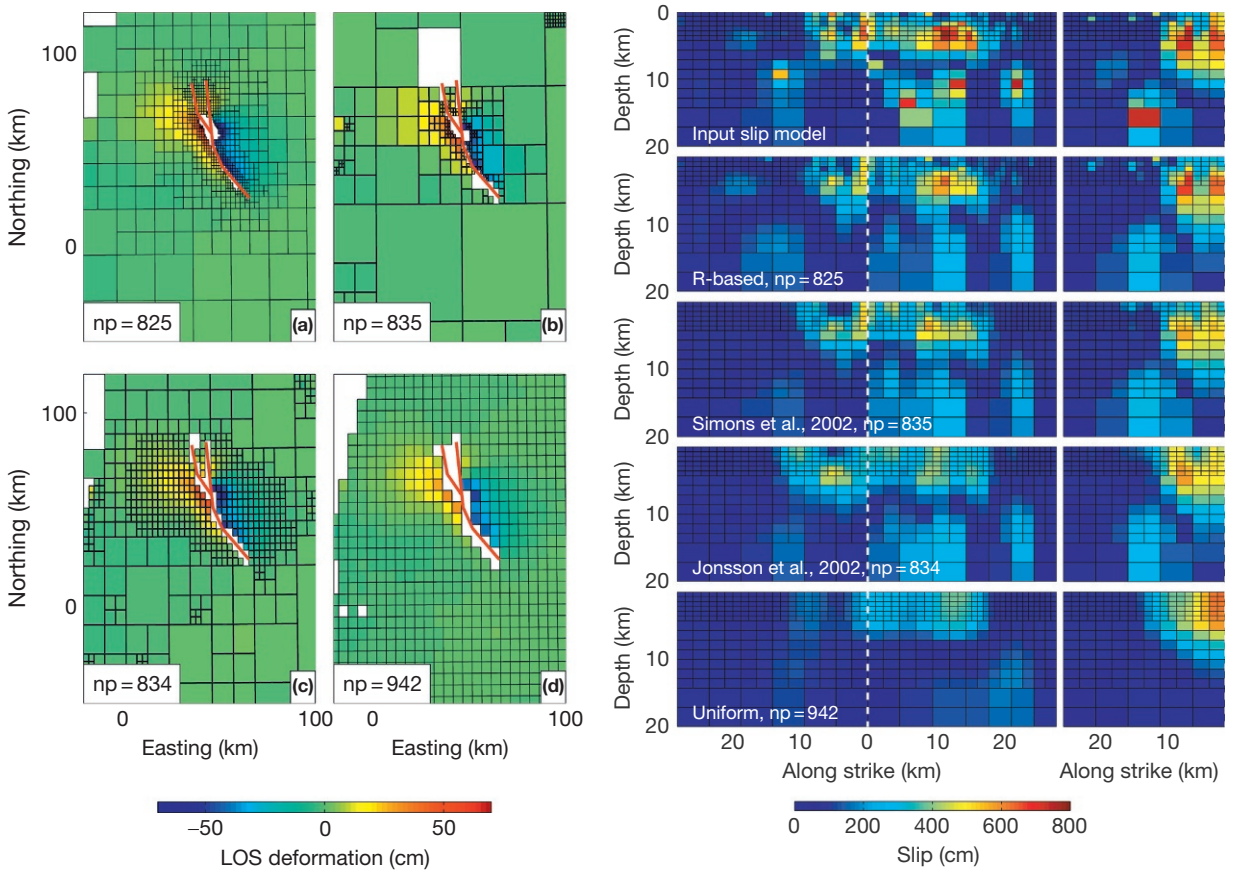


Figure 23 (Left) Examples of different interferogram sampling approaches. Results are shown using (a) the data resolution-based algorithm from Lohman and Simons (2005), (b) a variable sampling proportional to data curvature approach from Simons et al. (2002), (c) a variable sampling proportional to data gradient Jonsson et al. (2002), and (d) uniform sampling. (Right) Comparison of inferred slip distribution when using the different sampling approaches shown at the left. In each case, the total number of points retained is similar. The fault geometry is shown at the left by the red lines. The inferred slip is similar for all inversions, but the resolution of shallow slip features increases from bottom to top.

and near-fault decorrelation near shallow earthquake ruptures in the 1995 Kobe, Japan, earthquake (Rosen et al., 2000), the 1999 Hector Mine, California, earthquake (Simons et al., 2002), and the 2003 Bam, Iran, earthquake (Fielding et al., 2005). This later form of decorrelation has been found to clearly outline fault traces that slip in earthquakes and can serve as an obvious way to constrain surface fault geometry and should be used to guide post-earthquake field studies. In general, InSAR decorrelation images provide a synoptic view of the spatial extent (along strike and cross strike) of faulting that is not easily achieved from the ground.

Of more important potential societal benefit of decorrelation measurements comes from their use immediately after an earthquake, landslide, or eruption to map out regions of damage (e.g., Fielding et al., 2005). Frequently, existing communication systems are incapacitated by the event, and local inhabitants are not able to report the level of devastation. Rapidly made decorrelation maps produced during or soon after the event could be used to guide emergency response.

The usefulness of a decorrelation map is limited by 'uninteresting' sources of decorrelation not associated by the event. In particular, all forms of temporal decorrelation, especially if not monotonic, can mask the correlation image. Obvious examples include rain, snow fall, and vegetation changes occurring during the time span of the image pair. The most important ways to improve the usefulness of decorrelation measurements include rapid revisit times to limit the possibility of extraneous processes and the use of long wavelength InSAR (e.g., L-band as opposed to C-band) (Figure 24).

3.12.5 The Link Between Science and Mission Design

InSAR and other space geodetic techniques measure the displacement of the Earth's surface over time. The characteristics of the measurement – such things as its noise properties, sampling density in time and space, and area coverage rate – are tied to the specific implementation of the InSAR instrument and the mission characteristics. Clearly, there is a close link between the science that can be done with an InSAR mission and design of that mission. It is important for the research community to understand this link in order to accomplish their goals. A simple example of this would be avoiding the use of data from a mission that provides an image over an area once per month when the phenomenon of interest has changes that occur on timescales of days. In this section, we will describe the flow from scientifically driven measurement needs to the basic parameters of an InSAR mission. With such an understanding, it is then possible to characterize the performance of existing missions and productively discuss the design trades for future systems.

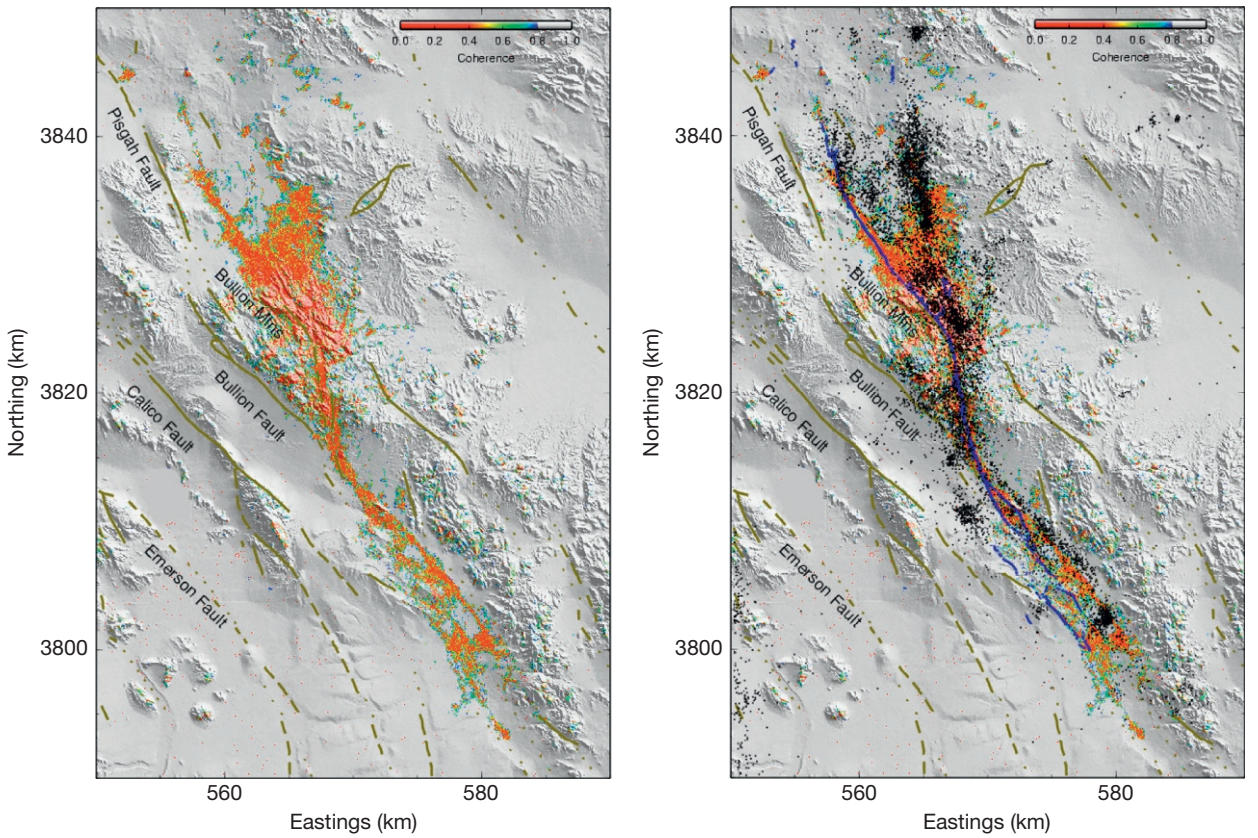


Figure 24 Interferometric decorrelation from the 1999 M_w 7.1 Hector Mine, California, earthquake. Only areas with correlation less than 0.8 are shown. Images are from the *C*-band ERS satellite and span 35 days. Same as left but with aftershocks and the fault trace as mapped in the field. Modified from Simons M, Fialko Y, and Rivera L (2002) Coseismic deformation from the 1999 M_w 7.1 Hector Mine, California, earthquake as inferred from InSAR and GPS observations. *Bulletin of the Seismological Society of America* 92: 1390–1402.

It is worth stating that InSAR measurements are a poor proxy for what scientists really would like to know about geophysical systems. The desired starting point to address larger geophysical questions – for example, what are the mechanics of earthquakes and how do fault systems interact? – would be measurements of the state of the crust, its pressure, temperature, and distribution of material properties throughout medium, to use as input to geophysical models. These *in situ* measurements at depth are impossible to obtain, so scientists model them through observations of the motions of the Earth's surface. So, it is important initially to understand the sensitivity relationship between model parameters and surface displacements. For example, if any reasonable change in a model parameter changes the modeled surface displacement by 1 mm, there is limited value in measuring deformation to only 1 cm accuracy.

So the questions arise: what can we reasonably expect to measure with an InSAR system, and conversely, what is required of that system in order to advance science? A repeat-pass InSAR system measures the range displacement of any image element through a differencing of the phase from one epoch to another. Using a time series of observations, a repeat-pass InSAR system measures a spatial distribution of range displacements in discrete time intervals. To use InSAR displacement measurements in geophysical models, the measurements must have adequate displacement accuracy, both absolute and relative, spatial resolution, spatial coverage, and temporal sampling. These requirements differ for each specific scientific investigation.

For a given system, we have seen that the accuracy of the range displacement measurement is determined by noise induced by the decorrelation of the radar echoes, by random phase delays introduced by propagation effects through the time-variable atmosphere, and by systematic knowledge uncertainties in the radar path delays and orbit. In addition, the incidence angle and azimuth angle of the observation can affect the accuracy of the measurement greatly: measuring a horizontal displacement with a system that looks steeply down toward nadir is not desirable.

As described earlier, decorrelation is composed of principally three components: thermal decorrelation is related to the noise level relative to the signal level of the radar system, geometric decorrelation is related to the arrangement of scatterers on the surface and how they change with differences in time or viewing geometry, and other decorrelation terms derive from noise (e.g., quantization, ambiguities, and side lobes) that is dependent on the signal level itself. There is a strong functional dependence of these decorrelation terms on system parameters, such as power, antenna size, and wavelength, and a great interplay among them.

The range resolution of a SAR system is inversely proportional to waveform bandwidth. The required range resolution is usually set by the scale size of the ground feature that is being mapped. For InSAR systems, however, there is a relationship between system bandwidth and desirable interferometric baselines. Finer resolution implies less decorrelation due to nonzero baselines. Thus, even if the final map resolution desired is only 100 m, one might require a system to have 10 m resolution in range so that the constraints on the repeat orbit accuracy are manageable.

The along-track resolution of a conventional strip map SAR system is equal to half the along-track length of the antenna, independent of range and wavelength. This is illustrated schematically in Figure 25, in which the antenna size is grossly exaggerated in size relative to the antenna beamwidth and the range, just so the salient characteristics can be viewed on a single page. The resolution along-track is determined by the spread in frequency content a given surface element experiences as the SAR beam illuminates it. Because of its beamwidth, the radar signal experiences Doppler shifts across its beam, such that the received echo contains a spectrum of information. The beamwidth is given by $\phi_D = \lambda/L$, where λ is the wavelength of the radar tone and L is the length of the antenna along-track. The frequency bandwidth associated with the Doppler shifts within this beamwidth is given by $\Delta f_{D,t} = 2v\phi_D/\lambda = 2v/L$. In spatial coordinates, the spatial frequency bandwidth is $\Delta f_{D,x} = 2/L$, which implies a resolution of $L/2$. Thus, while wavelength and antenna length determine the beamwidth of the antenna, the spatial frequency content is not dependent on the wavelength in SAR.

For ScanSAR systems, the resolution is determined by the length of the burst of pulses in a given scan, which is generally related to the number of ScanSAR beams. For spotlight SAR systems, the resolution is determined by the length of time over which the observation is made.

In addition to the thermal noise present in a radar system, there are a number of noise sources that play a significant role in the design of a radar and the accuracy one can expect to achieve for InSAR displacements. SAR systems point to an angle off nadir to avoid echoes from both sides of the nadir track: such echoes would be ambiguously combined in the SAR receiver and could not be distinguished from each other. Even with off-nadir pointing, the transmitted and received energy cannot be completely localized in time as the signal spreads throughout the illumination area, resulting in a wide range of time over which a given echo can return (and even some energy from the opposite side of nadir in some cases). Because the radar transmits pulses of energy over 1000 times per second for typical spaceborne systems in order to properly sample the Doppler spectrum, it can often occur that energy from time intervals outside the area of interest defined by this sampling rate, for example, from a previous or later pulse, arrives at the receiving antenna at the same time as the desired echo energy. These corrupting echoes, generally occurring at much lower amplitude, are called range ambiguities. The magnitude of these are controlled by the pulse rate – generally lower pulse rate allows more time to collect all echoes from a pulse – and by shaping the illumination area by manipulating the antenna pattern.

On the other hand, one cannot lower the pulse rate below the point where the along-track pulses become undersampled. As we have seen, to create a narrow Doppler spectrum, we desire a long antenna in the along-track direction. Physical constraints on the size of the flight system, as well as a common desire for reasonably high resolution, limit this size and therefore limit the minimum pulse rate. The illumination pattern in the along-track direction also has extent beyond the nominal beamwidth (antenna sidelobes), so the pulse sampling rate will naturally cause aliasing of some energy from beyond the along-track antenna main beam extent.

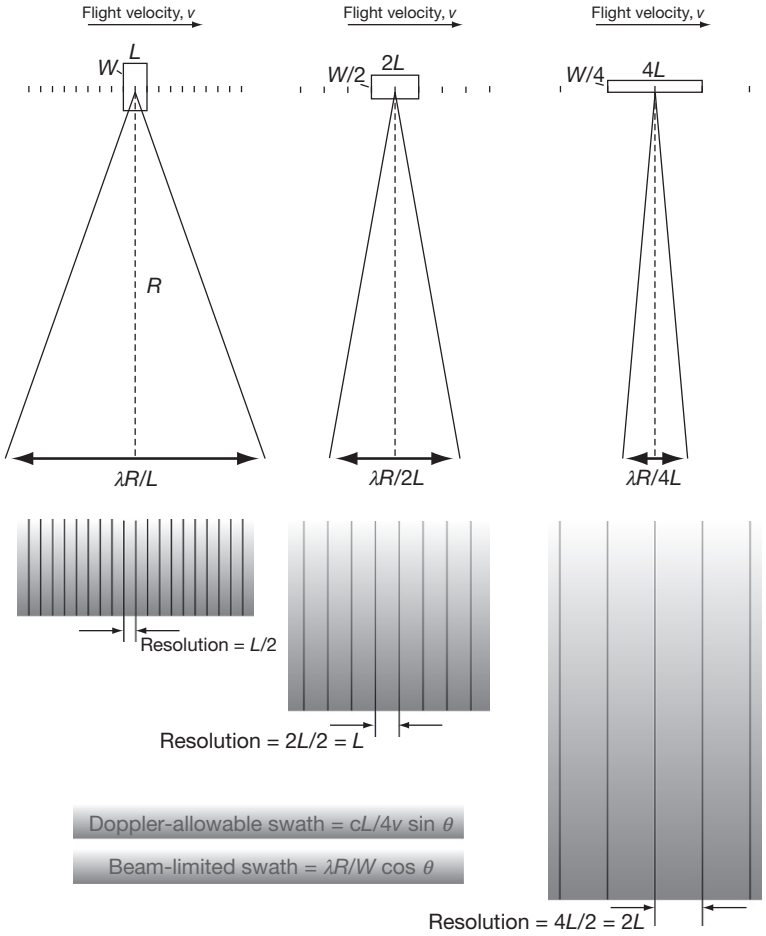


Figure 25 Illustrating the trade-off between swath extent (coverage) and resolution for strip mode SAR systems. In order to image a wide swath in strip mode, the antenna must be long in the along-track dimension. However, the resolution along-track is half the antenna dimension, so the resolution is degraded as the swath width increases.

To first order, then, if a wide swath is desired, then a low pulse rate must be chosen to allow enough time between pulses for the received echo to be unambiguously acquired. This then requires that the Doppler bandwidth and hence the antenna beamwidth be narrow, which then mandates a long antenna in the along-track direction. Furthermore, the antenna must be limited in size in the elevation direction to create a wide enough beam to receive energy from the wide swath of interest.

For a particular spacecraft altitude, the swath size is determined by these ambiguity constraints for most practical spaceborne systems. As the radar antenna of fixed beamwidth is pointed at greater distances off nadir, the swath illuminated on the ground becomes broader from projection effects, but the usable swath extent is usually narrower because of ambiguities. This then means that the antenna must be larger in the elevation dimension to limit the swath to maintain performance.

These effects then influence mission design for an InSAR system. Scientists studying deformation want to be able to observe any point on the Earth at some regular interval. They also would like rapid repeat coverage to be able to track rapid changes of the Earth. Suppose the mission requirement is to repeat an orbit every 12 days. For exact repeat, this requires an integer number of orbits in this time. There is a 12-day repeat polar orbit at about 690 km altitude that contains 180 orbits in 12 days. This then implies that the separation of the orbit tracks at the equator will be $2\pi 6378/180 \approx 220$ km. Thus, the SAR must be able to cover 220 km of swath either all at once with a very wide swath from a very long and skinny antenna (giving low resolution) or using multiple smaller beams with smaller swaths covering different angles off nadir at different times. The previously described ScanSAR or TOPS SAR systems, where the radar sends a collection of pulses illuminating one subswath, then electronically steers the antenna to the next subswath off nadir, and sends the next collection of pulses and so on with multiple beams, allow full coverage of the wide swath in one pass, at the expense, again, of resolution. There are also new techniques and technologies being developed that allow an antenna to be partitioned into multiple subantennas, with transmission of multiple signals simultaneously, and similarly simultaneous multiple reception, opening the trade space to broad coverage while preserving resolution.

These design space possibilities – frequency, resolution, antenna size and functionality, orbit altitude and control, system power, viewing angle, repeat period, and observation

modality – are the playground of scientists, working with system designers, to optimize a mission to capture meaningful geophysical signals in the presence of the noise sources that are present in radar measurements. Spacefaring nations are increasingly relying on SAR and InSAR for scientific discovery and monitoring, with the trend moving away from large multimode systems to simpler instruments that do a few things well. One of these missions will no doubt be an InSAR mission, a true geodetic instrument in the model of GPS, but with global reach and orders of magnitude denser coverage of the Earth.

Appendix A

A.1 Radar Imaging

The power of radar interferometry for geodetics stems from its high-resolution images that are generated from a controlled coherent light source. The coherence of the measurements ensures that the phase associated with each complex image element contains both the round-trip geometric path length from the radar sensor to the surface and back and the ensemble phase associated with the coherent summation of scattering within the image element. The imaging process is quite involved and is described in detail in many other places, for example, Elachi (1988). However, some basic description of the process will frame the discussion of interferometry. This appendix is a compilation of the basics from a variety of sources and is designed to cover all the essentials for understanding radar images, and their phase, in one convenient narrative.

The most common radar waveform in civilian spaceborne radars is *chirp* coding:

$$f(t) = A\cos\left(\omega_0 t + \frac{B_r}{2\tau_p}t^2\right)\mathrm{rect}\left(\frac{t}{\tau_p}\right) \quad [32]$$

This waveform is transmitted typically over 1000 times per second and its echo is received, digitized, and transmitted to a recorder. The presence of a point target on the ground surrounded by a surface with no inherent reflectivity causes the radar system to respond with the following received signal:

$$r(t,R) = p(R)g(R)\cos\left(\omega_0(t-2R/c) + \frac{1}{2}\frac{B_r}{\tau_p}(t-2R/c)^2\right)\mathrm{rect}\left(\frac{t-2R/c}{\tau_p}\right) \quad [33]$$

where, for now, $p(R)$ and $g(R)$ are generic functions to represent $1/R^2$ loss in power and backscatter at R, respectively. They will be defined with greater specificity shortly. Note if we think of the radar as a system, the point target can be thought of as a delta-function input, implying that $r(t, R)$ would be the system impulse response.

If numerous point targets are arranged in range, then the total received signal at any given time will be the sum of point target returns over these ranges:

$$(t) = \int_R p(R)g(R)\cos\left(\omega_0(t-2R/c) + \frac{1}{2}\frac{B_r}{\tau_p}(t-2R/c)^2\right)\mathrm{rect}\left(\frac{t-2R/c}{\tau_p}\right)\mathrm{d}R \quad [34]$$

This equation states that the received signal is the convolution of the ground backscatter, modulated by the geometric decay in power with range, with the transmitted chirp. Viewing the transmitted chirp as the radar system impulse response, this equation is a typical expression for the output of a linear system driven by $p(R)g(R)$.

The response of the radar to the presence of a line of scatterers in range, represented in its complex-valued form $r_z(t)$ is

$$r_z(t) = K\int_{t'} p(t')g(t')\mathrm{e}^{\mathrm{j}\left(\omega_0(t-t') + \frac{1}{2}\frac{B_r}{\tau_p}(t-t')^2\right)}\mathrm{rect}\left(\frac{t-t'}{\tau_p}\right)\mathrm{d}t' \quad [35]$$

where K is an arbitrary constant and $t' = 2R/c$. The convolution of the surface with the chirp smears out the information from any given scatterer over the extent τ_p. The goal is to recover the original $g(t')$ from $r_z(t)$.

A.1.1 Matched filtering

Suppose

$$p(t')g(t') = g_0\mathrm{e}^{\mathrm{j}(\omega_0 t' - 2kR_0)}\delta(t'-t_0) \quad [36]$$

where $R_0 = ct_0/2$, implying there is a single point scatterer on the ground at range R_0. In the radar receiver, normally one of the first hardware functions is to essentially remove the rapidly varying phase term $\exp(\mathrm{j}\omega_0 t')$ by a process called heterodyning. The received signal is multiplied by a reference signal $\exp(-\mathrm{j}\omega_0' t')$, with $\omega_0 \approx \omega_0'$ so the narrowband received signal (the bandwidth is $B \ll \omega_0$) is near zero frequency, in the so-called video band. (In practice, there can be several stages of filtering and heterodyning to achieve the video signal, and often, it is arranged so the spectrum of the real signal has its positive and negative frequencies symmetrically arranged around zero, each centered at $B/2$ and $-B/2$, respectively.) For our purposes, we let $\omega_0 \approx \omega_0'$, so that

$$p(t')g(t') = g_0\mathrm{e}^{-\mathrm{j}2kR_0}\delta(t'-t_0) \quad [37]$$

with the understanding that the 'carrier' has been removed. Then

$$\begin{aligned} r_{z\delta}(t) &= g_0\mathrm{e}^{-\mathrm{j}2kR_0}\mathrm{e}^{\mathrm{j}\frac{1}{2}\frac{B_r}{\tau_p}(t-t_0)^2}\mathrm{rect}\left(\frac{t-t_0}{\tau_p}\right) \\ &= g_0\mathrm{e}^{-\mathrm{j}2kR_0}h(t-t_0) \end{aligned} \quad [38]$$

Since $r_{z\delta}$ is the response of the radar to an impulsive 'signal' source (i.e., the ground), we recognize

$$h(t) = \mathrm{e}^{\mathrm{j}\frac{1}{2}\frac{B_r}{\tau_p}t^2}\mathrm{rect}\left(\frac{t}{\tau_p}\right) \quad [39]$$

as the impulse response of the radar system. For this system, h is just the transmitted signal. The Fourier transform of h is known as the system transfer function $H(\omega)$. One of the properties of Fourier transforms is that the convolution of functions in the time domain is equivalent to the product of functions in the frequency domain

$$r_z(t) = \int g(t')h(t-t')\mathrm{d}t' \leftrightarrow R_z(\omega) = G(\omega)H(\omega) \quad [40]$$

Since we are interested in recovering $g()$ from $r_z()$, it is clear that if we can multiply $R_z(\omega)$ by the reciprocal of $H(\omega)$, the inverse Fourier transform will retrieve $g()$. It turns out that $H(\omega)$ is of the form

$$H(\omega) \approx e^{j\kappa\omega^2} \quad [41]$$

which implies $R_z(\omega)H^*(\omega) = G(\omega)$, that is, we can filter r_z with the conjugate spectrum of the impulse response to recover g.

This specific result is a special case from matched-filter theory, where an optimal filter is designed to best detect a transmitted waveform in the presence of noise (Figure 26). The matched filter of a signal $s(t)$ has an impulse response

$$h_m(t) = s^*(-t) \quad [42]$$

In our case, $s()$ is what we defined as radar impulse response $h()$, so $h_m(t) = h^*(-t)$. Thus, the recovered, or compressed, signal r_{zc} for our delta function is given by

$$r_{zc}(t) = \int r_z(t')h_m(t-t')\,dt' \quad [43]$$

$$\begin{aligned} r_{zc\delta}(t) &= g_0 e^{-j2kR_0}\int h(t'-t_0)h_m(t-t')dt' \\ &= g_0 e^{-j2kR_0}\int h(t'-t_0)h^*(t'-t)dt' \end{aligned} \quad [44]$$

So we see that convolution with the matched filter is equivalent to correlation with the conjugate of the function itself. This makes some sense intuitively, in that as we slide the filter along our received signal and integrate, they will match best and give the highest integral when they are aligned with lag $t = t_0$ and when all phase terms are canceled making the integrand always positive. Since the Fourier transform of $h^*(-t)$ is just $H^*(\omega)$, we see that the result above considering the specific form of h in eqn [41] is actually a more general result. Namely, we can in general filter r_z with the conjugate spectrum of H to retrieve g, or equivalently,

$$r_{zc}(t) = \int r_z(t')e^{-j\frac{1}{2}\frac{B_r}{\tau_p}(t'-t)^2}\,\mathrm{rect}\left(\frac{t'-t}{\tau_p}\right)dt' \quad [45]$$

Expanding eqn [44], we have

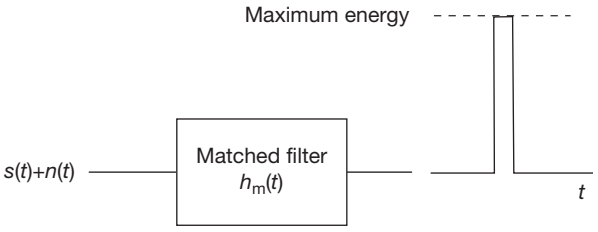


Figure 26 Illustration of the matched-filter concept, where a signal in noise is optimally detected by passing it through a filter matched to its characteristics. In our case, $s(t)$ is what we called the impulse response of the radar system $h()$.

$$\begin{aligned} r_{zc\delta}(t) = g_0 e^{-j2kR_0}\int e^{j\frac{1}{2}\frac{B_r}{\tau_p}(t'-t_0)^2}\,\mathrm{rect}\left(\frac{t'-t_0}{\tau_p}\right)e^{-j\frac{1}{2}\frac{B_r}{\tau_p}(t'-t)^2} \\ \mathrm{rect}\left(\frac{t'-t}{\tau_p}\right)dt' \end{aligned} \quad [46]$$

If we ignore the rect functions and assume infinite integration limits, then the integration becomes fairly straightforward because the quadratic phase terms involving the integration variable cancel:

$$\begin{aligned} r_{zc\delta}(t) &= g_0 e^{-j2kR_0} e^{j\frac{1}{2}\frac{B_r}{\tau_p}(t_0^2-t^2)}\int e^{-j\frac{B_r}{\tau_p}(t't_0)} e^{j\frac{B_r}{\tau_p}(t't)}\,dt' \\ &= g_0 e^{-j2kR_0} e^{j\frac{1}{2}\frac{B_r}{\tau_p}(t_0^2-t^2)}\int e^{j\frac{B_r}{\tau_p}t'(t-t_0)}\,dt' \\ &= g_0 e^{-j2kR_0} e^{-j\frac{1}{2}\frac{B_r}{\tau_p}(t^2-t_0^2)}\,\delta(t-t_0) \end{aligned} \quad [47]$$

So we see that we exactly recover the delta function with an infinite bandwidth signal and impulse response. If we were to leave the impulse response bandwidth infinite but apply the windowing of the received signal, then the integration limits in the preceding text are limited to the pulse extent τ_p, and we obtain the usual sinc result:

$$r_{zc\delta}(t) = g_0 e^{-j2kR_0} e^{-j\frac{1}{2}\frac{B_r}{\tau_p}(t^2-t_0^2)}\frac{\sin\pi B(t-t_0)}{\pi B(t-t_0)} \quad [48]$$

where $B = B_r/2\pi$ (Figure 27). The first null is at the expected location of $1/B$, defining the resolution of the compressed signal.

When both rect functions are considered in the integral, the quadratic phase term cancels and the sinc function is modified. The integration is carried out in three regimes as shown in Figure 28. Using these limits,

$$r_{zc\delta}(t) = g_0 e^{-j2kR_0}\,\mathrm{rect}\left(\frac{t-t_0}{\tau_p}\right)\frac{\sin\pi B(t-t_0)(\tau_p - |t-t_0|)}{\pi B(t-t_0)} \quad [49]$$

When $t - t_0 \ll \tau_p$, this reduces to the usual expression without the quadratic phase:

$$r_{zc\delta}(t) \approx g_0 e^{-j2kR_0}\frac{\sin\pi B(t-t_0)}{\pi B(t-t_0)} \quad [50]$$

A.1.2 Radar signal return from a three-dimensional surface

We examine what comprises $p()$ and $g()$ in eqn [35] in an imaging radar system looking at the Earth, where we do not necessarily have a neat line of scatterers arranged in range. Instead, we have a spatial distribution of scatterers within the beam on a three-dimensional topographically modulated surface (refer back to Figure 2 for the SAR imaging geometry). The receive echo will integrate over the azimuth direction and the

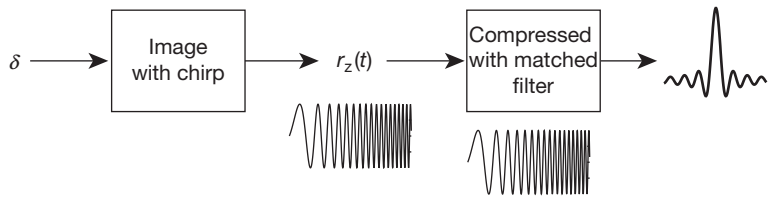


Figure 27 Illustration of the matched filter for the delta-function scene and linear FM chirp. The compressed signal is close to a sinc function.

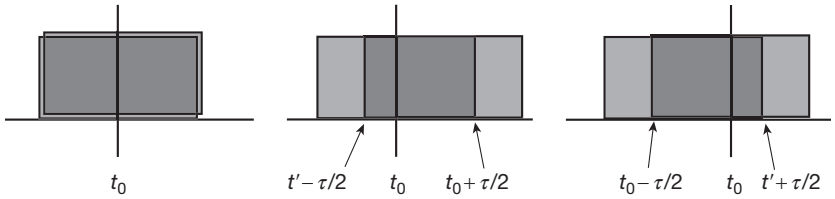


Figure 28 Showing the three regimes of integration as the windowed received signal and matched-filter lag.

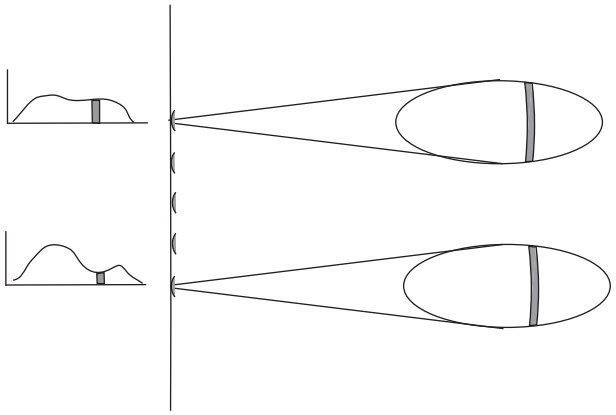


Figure 29 Radar imaging as viewed looking down at the Earth from above the sensor. The sweep of the pulse across the radar beam leads to a return echo trace for each pulse as depicted on the left. Periodically, a pulse is sent and a trace recorded. The collection of pulses, that is, the along-track history of range profiles, is therefore a two-dimensional representation of the surface, albeit poorly resolved. There is considerable redundant information for any given surface element in the pulses because the beam samples each point many times as is travels along-track (e.g., pulses every 5 m observed over a beam extent of 5 km). Through matched-filtering techniques, a fine-resolution image of the surface can be generated.

range pulse at any instant. For a real aperture radar, that is one for which the resolution in azimuth is dictated by the beamwidth of the antenna, the point target concept extends to an integrated backscatter value over azimuth $g'(R)$ and the same arguments and equations for the simple point targets described in the preceding text obtained.

Figure 29 shows the imaging scenario looking down on the Earth from above the sensor. This more clearly shows that the received echo energy at any given instant within the pulse will effectively be the integrated return from every scatterer within the instantaneous pulse extent on the ground in range and azimuth, weighted by the transmitted chirp waveform. Because the chirp signal is extended in range, the response from any given scatterer is also extended in range. Variations in range relative to a nominal flat surface within the beam induced by topography or surface relief will lead to geometric distortion in the range trace: the measurement is fundamentally an integration (convolution) over phase fronts (i.e., constant range).

Being more specific about the function $g()$, we define the geometry of the phase front incident on a elemental surface with reference to Figure 30. The surface element is like the facet model, bounded in range by a distance $\triangle R$, and presents a cross-sectional area normal to the incident wave of $\mathrm{d}a\,\mathrm{d}z$, where $\mathrm{d}a = R\,\mathrm{d}\phi$ and $\mathrm{d}z = \triangle R/\tan\theta$, where θ is the look angle. With this definition, we can define the relationship between the normalized back scattering cross subsection σ_0 and the reflectivity normal to the look direction.

$$\left|\Gamma(a,R)\right|^2 \mathrm{d}a\,\mathrm{d}z = \frac{\sigma_0(a,R)}{\sin\theta}\mathrm{d}a\,\Delta R \qquad [51]$$

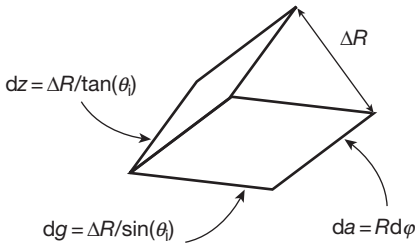


Figure 30 Depiction of a cell in range defined by two phase fronts separated by range distance ΔR. This could be the transmit pulse length in range units, for example.

This equation defines Γ. In general, the reflectivity is a complex quantity, taking into account phases from random scatterers and differential path delays within the elemental area.

The received field from this elemental area is then

$$r(t) = \frac{P_T^{1/2} G\lambda}{(4\pi)^{3/2} R^2}\Gamma(R,\phi)R\,\mathrm{d}\phi\,\mathrm{d}z\,p(\theta(R),\phi)\mathrm{e}^{-2\mathrm{j}kR}\mathrm{e}^{\mathrm{j}\frac{1}{2}\frac{B_r}{\tau_p}(t-2R/c)^2}$$
$$\mathrm{rect}\left(\frac{t-2R/c}{\tau_p}\right) \qquad [52]$$

At every instant $t-2R/c$, all the signals that arrive at the receiver add up. We have constructed the elemental area on the phase front such that an integration over these elements is at fixed $2R/c$. Thus, we can define the integration over the phase front to be a function of time that is a generalization of the range-only function $g()$ studied earlier:

$$g_\Gamma(R) = \mathrm{e}^{-\mathrm{j}2kR}\iint p(\theta(R),\phi)\Gamma(R,\phi)R\,\mathrm{d}\phi\,\mathrm{d}z \qquad [53]$$

If the scatterers were all arranged on a single phase front at range R, then the received signal would be

$$r(t) = \frac{P_\mathrm{T}^{1/2}G\lambda}{(4\pi)^{3/2}R^2}g_\Gamma(R)\mathrm{e}^{\mathrm{j}\frac{1}{2}\frac{B_\tau}{\tau_\mathrm{p}}(t-2R/c)^2}\,\mathrm{rect}\left(\frac{t-2R/c}{\tau_\mathrm{p}}\right) \qquad [54]$$

For scatterers arranged over the entirety of the surface, we have an additional integration over all phase fronts.

$$r(t) = \frac{P_\mathrm{T}^{1/2}G\lambda}{(4\pi)^{3/2}R^2}\int g_\Gamma(R)\mathrm{e}^{\mathrm{j}\frac{1}{2}\frac{B_\tau}{\tau_\mathrm{p}}(t-2R/c)^2}\,\mathrm{rect}\left(\frac{t-2R/c}{\tau_\mathrm{p}}\right)\mathrm{d}R \qquad (55)$$

This is a convolution of the transmitted pulse with the integrated phase front. The range compression process (matched filtering) described in the preceding text returns to us a record of the total return from phase fronts resolved in time to $1/B$ or in range to $c/2B$. The next step is to exploit the redundancy in the range signal as a function of pulse number to build resolution in the along-track dimension.

A.1.3 Azimuth signal and aperture synthesis

Real aperture radars are very useful for a variety of applications, generally those which require detection of a bright object in a broad region as is common in military applications, or where a low-resolution averaged measure of surface roughness or dielectric is sufficient. Oceans and ice sheets are generally featureless, so high resolution is not often needed. Large-area estimates of backscatter can be interpreted as wind speed, for example, on global scales, with the sensor dealing with only modest amounts of data.

Often, however, especially for land applications, fine-resolution images are important because the natural spatial variability of the surface is rapid. So it would be very nice if we could overcome the azimuth beam limit on resolution without building an enormously long antenna. It turns out that it is possible using the techniques of aperture synthesis.

Returning momentarily to a point target on the ground, consider a train of pulses transmitted as a spacecraft travels along a straight line path (Figure 31) above a flat Earth. The pulse is transmitted and subsequently received at a slightly later time, but for the purpose of understanding the concepts, we can imagine 'stop-and-shoot' model where the radar to target to radar distance is $2R_i$ for a given pulse i. The transmitted and received pulses are shown schematically in the figure, illustrating the changing range from pulse to pulse.

In this geometry, $R_i = \sqrt{R_0^2 + (x_i - x_0)^2}$, $x_i = vt_i$, and $t_i = i\mathrm{PRI} + t_0$, where R_0 is the shortest range from the flight track to the target, also called 'broadside,' and x_0 is the azimuth at that location. Thus, the phase factor we have been carrying along for a return from a broadside pulse, $\exp(-\mathrm{j}2kR_0)$, must now be updated from pulse to pulse. After demodulation and range compression, the received signal from a point target for pulse i is

$$r_{\mathrm{zc}\delta,x_i}(t) = g_0 p(x_i - x_0)\mathrm{e}^{-\mathrm{j}2kR_i(x_i)}\mathrm{sinc}\{\pi B(t-t_i)\} \qquad [56]$$

where $t_i = 2R_i/c$. This is the range compressed form eqn [50] with an adjusted range term and an explicit function $p()$ denoting the antenna pattern of the radar antenna. Writing this entirely in terms of range and azimuth coordinates gives

$$r_{\mathrm{zc}\delta,x_i}(R_\mathrm{f}) = g_0 p(x_i - x_0)\mathrm{e}^{-\mathrm{j}2kR_i(x_i)}\mathrm{sinc}\left\{\frac{\pi}{\Delta R}(R_\mathrm{f} - R_i)\right\} \qquad [57]$$

where R_f is the range variable, also called 'fast time' because it represents the time variation of the signal across a single pulse, and x_i is the azimuth location of a given pulse, also called 'slow time' since it characterizes the slower variation of the range trace from pulse to pulse.

If the transmitting antenna is length L in azimuth, then the target will be illuminated for an azimuth distance (3-dB) of

$$X_\mathrm{ill} = \frac{\lambda}{L}R_0 \qquad [58]$$

The echo field from the point target will be sampled $N = X_\mathrm{ill}/\Delta x_\mathrm{p}$ times while it is in the beam, where $\Delta x_\mathrm{p} = v$ PRI is the pulse spacing on the flight track (Figure 32). The field is sampled effectively by a linear array of identical antennas spaced by Δx_p. It is shown in the homework (and any number of books) that the antenna pattern of a linear array of this sort is the product of the antenna pattern of the physical antenna

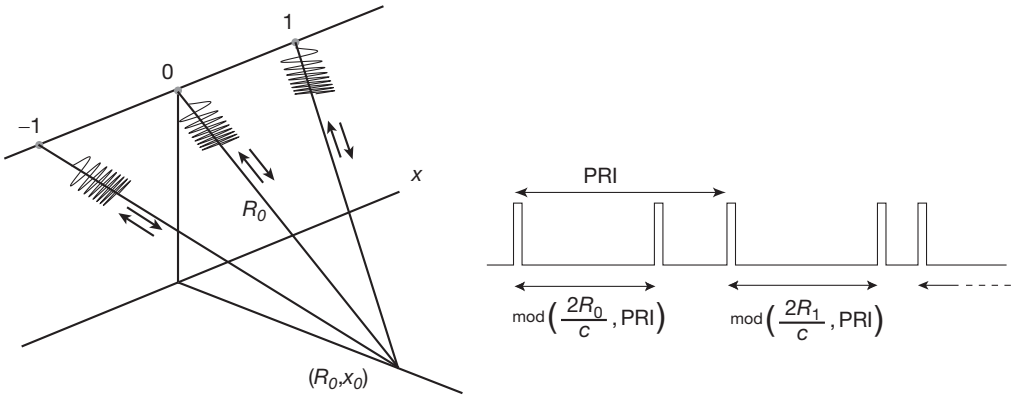


Figure 31 Illustration of range history in azimuth of the pulsed radar. Ignoring the finite speed of light, the radar can be thought of as transmitting and receiving from each point on orbit. Because of the finite extent of the azimuth beam, a single point target on the ground will be seen in a succession of received pulse echoes, with decreasing to increasing range location as the spacecraft flies by.

and the array factor. The physical antenna is small leading to a broad sinc-like beam in azimuth. The array factor is also sinc-like but is much narrower:

$$\mathrm{AF} = \frac{\sin\left(\frac{N}{2}k\Delta x_\mathrm{p}\sin\theta\right)}{\sin\left(\frac{1}{2}k\Delta x_\mathrm{p}\sin\theta\right)} \quad [59]$$

In the limit as $\triangle x_\mathrm{p} \to 0$, and for small θ,

$$\mathrm{AF} \to N\frac{\sin\left(\frac{k}{2}X_\mathrm{ill}\theta\right)}{\frac{k}{2}X_\mathrm{ill}\theta} \quad [60]$$

which has the expected 3-dB beamwidth of $\theta_{3\text{-dB}} = \lambda/X_\mathrm{ill}$. Thus, the sampled linear array in space serves as a very long antenna of length X_ill. Without careful deliberation, we might conclude that the resolution in azimuth we could achieve is given by its effective 3-dB beam extent $\triangle X = \theta_{3\text{-dB}}R_0 = L$; however, we must take into account the two-way propagation of the radar signal. The earlier arguments are correct for an antenna synthesized to illuminate an area at range R_0. However, since $2R_0$ is the actual distance propagated, we can think of the beam extent for illumination as actually at twice the distance, that is, the phase variation across the aperture is as if the antenna were at twice the distance. Thus,

$$X_\mathrm{ill,\,eff} = \frac{\lambda}{L}2R_0 \quad [61]$$

and the resolution allowed by an antenna of this length is

$$\Delta X = L/2 \quad [62]$$

This remarkable result states that the resolution of a synthetic aperture system is independent of range and velocity and is just half the physical antenna length. As the range is increased, the synthetic aperture increases in length and the angular extent of synthetic beam narrows in proportion to maintain fixed resolution. If the physical antenna decreases in size for a fixed range, the illuminated area increases, increasing the synthetic aperture and narrowing the synthetic beamwidth. Since the range is fixed, the resolution becomes finer in proportion to the reduction in L.

Consider R_i as a continuous variable of the azimuth coordinate

$$\begin{aligned} R(x) &= \sqrt{R_0^2 + (x - x_0)^2} \\ &\approx R_0\left(1 + \frac{(x-x_0)^2}{2R_0^2}\right) \\ &\approx R_0 + \frac{(x-x_0)^2}{2R_0} \end{aligned} \quad [63]$$

The constant phase term $\exp(-\mathrm{j}2kR_0)$ is very difficult to measure absolutely (billions of cycles), so for now, we will ignore it and focus on the spatially varying component of the phase

$$\mathrm{e}^{-\mathrm{j}2k(x-x_0)^2/2R_0} = \mathrm{e}^{-\mathrm{j}\pi(x-x_0)^2/F^2} \quad [64]$$

where $F^2 = \lambda R_0/2$. F is known as the Fresnel zone and it is the distance along the synthetic array where the wave front is within $\pi/4$ radians of phase (see Figure 33). Typical Fresnel zone size is on the order of 100–200 m. Since the array is coherent over this length, points simply summed together in azimuth will add coherently and form a meaningful, albeit low-resolution representation of the surface in azimuth:

$$\pi\frac{(x-x_0)^2}{F^2} < \frac{\pi}{4} \to x - x_0 < \frac{F}{2} \quad [65]$$

The range compressed pulse history can now be written

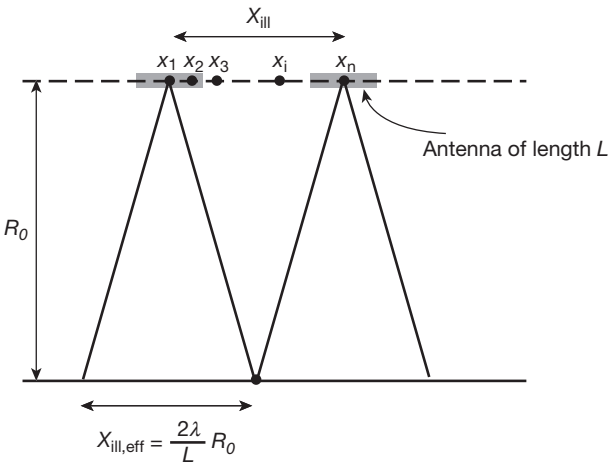


Figure 32 Point target illuminated by an antenna of length L. The 3-dB extent of the beam is $\lambda R_0/L$, but because two-way propagation introduces a doubling of the range and consequently the phase, the effective extent of illumination is $2\lambda R_0/L$.

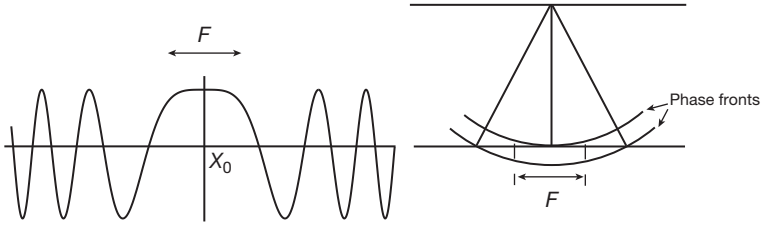


Figure 33 Illustration of the Fresnel zone, where the phase from an incident wave front varies by less than $\pi/4$ radians across the zone (right). Symmetric chirp at the left is the real part of the azimuth chirped phasor. It is nearly constant over the Fresnel zone. The synthetic array is coherent over this length and points simply summed in azimuth will add coherently and will be resolved at this length.

$$r_{zc\delta}(x_i, R_f) = g_0 p(x_i - x_0) e^{-j2kR_0} e^{-j\pi(x_i - x_0)^2/F^2} \operatorname{sinc}\left\{\frac{\pi}{\Delta R}\left[R_f - \left(R_0 - \frac{(x_i - x_0)^2}{2F^2}\right)\right]\right\} \quad [66]$$

where we explicitly call out the functional dependence on x_i in the argument to $r_{zc\delta}$.

The form of the azimuth signal is rather similar to the range signal before compression. Using the same approach to recover the point target in azimuth, we can define a matched filter in azimuth that is a similar conjugate chirp signal:

$$h_{am}(x) = h_a^*(-x) = p(-x) e^{j\pi x^2/F^2} \quad [67]$$

The doubly compressed point target signal is then

$$r_{zcc\delta}(x, R) = g_0 e^{-j2kR_0} \operatorname{sinc}\left\{\frac{\pi}{\Delta R}[R - R_0]\right\} \cdot \int p(x' - x_0) e^{-j\pi(x' - x_0)^2/F^2} p(x' - x) e^{j\pi(x' - x)^2/F^2} dx' \quad [68]$$

where we ignored the x-dependent term in the range sinc function because processors generally take care of this term before performing the matched filter in azimuth. If we allow $p()$ to be a rect function of extent X_{ill} as we did for range, then we will obtain exactly the same expression as we did for range, and the doubly compressed point target signal becomes

$$r_{zcc\delta}(x, R) = g_0' e^{-j2kR_0} \operatorname{sinc}\left\{\frac{\pi}{\Delta R}[R - R_0]\right\} \operatorname{sinc}\left\{\frac{\pi(x - x_0)X_{ill}}{F^2}\right\} \quad [69]$$

Since in this case we explicitly took into account the two-way propagation in the phase, X_{ill} is the actual illuminated extent, not the double-length effective extent used in the heuristic argument earlier. The first null of the azimuth response is where

$$x - x_0 = \frac{F^2}{X_{ill}} = \frac{\lambda R_0/2}{\lambda R_0/L} = \frac{L}{2} \quad [70]$$

We take this as the 3-dB resolution in azimuth, and we see it is identical to the heuristically derived resolution.

Expressed in terms of the range resolution $\triangle R = c/2B$ and azimuth resolution $\triangle X = L/2$, the overall impulse response of the radar system and matched-filtering operations is

$$r_{zcc\delta}(x, R) = g_0' e^{-j2kR_0} \operatorname{sinc}\left\{\frac{\pi}{\Delta R}[R - R_0]\right\} \operatorname{sinc}\left\{\frac{\pi}{\Delta X}[x - x_0]\right\} \quad [71]$$

For a surface described by a general reflectivity function in range and azimuth $\Gamma(x, R)$ related to the surface backscatter cross subsection through eqn [51], the matched-filter response will be the linear convolution with the impulse response in eqn [71]:

$$r_{zcc}(x, R) = \int e^{-j2kR'} \Gamma(x', R') \operatorname{sinc}\left\{\frac{\pi}{\Delta R}[R - R']\right\} \operatorname{sinc}\left\{\frac{\pi}{\Delta X}[x - x']\right\} dx' dR' \quad [72]$$

A.1.4 Range-Doppler images

SAR images generated in this style of filtering are often called 'range-Doppler' images. The quadratic variation of the phase of a point target with time along-track leads to a linear frequency variation.

$$\phi(t) = -\frac{\pi}{F^2} v^2 (t - t_0)^2 \quad [73]$$

$$\omega(t) = -2\pi \frac{v^2}{F^2}(t - t_0) \quad [74]$$

$$f_D(t) = -\frac{v^2}{F^2}(t - t_0) \quad [75]$$

Note the Doppler frequency is positive as the sensor approaches the target ($t < t_0$). The Doppler bandwidth is the totality of frequency content for any target. This is limited by the beamwidth in azimuth, $T_{ill} = X_{ill}/v$, giving

$$f_{D,ill} = \frac{2}{\lambda R} v^2 \frac{X_{ill}}{v} = \frac{2v}{L} \quad [76]$$

This is in time units (Hz). In spatial units, $f_{Dx'ill} = 2/L$, which is just the reciprocal of the spatial resolution.

Another way of describing this Doppler bandwidth is through the Doppler equation $f_D = -2\vec{v} \cdot \hat{l}/\lambda$, where $\vec{v}$ is the vector velocity of the sensor and $\hat{l}$ is the direction from the sensor to a point on the ground. At any instant of time, there is a one-to-one correspondence between the azimuth position of a target and the Doppler frequency (Figure 34). The antenna beam illuminates an area on the ground, with each target at a given angle off boresight having a unique Doppler frequency. The Doppler frequency can be written as $f_D = -2v \cos\theta_{az}/\lambda$, where θ_{az} is the angle formed by the velocity vector $\vec{v}$ and the look vector $\hat{l}$. Defining $\theta_{b.s.}$ as the angle measured relative to broadside (i.e., perpendicular to the velocity vector) and assuming that the antenna's boresight is oriented at broadside, $f_D = -2v \sin\theta_{b.s.}/\lambda$. The total Doppler bandwidth at that instant is given by the Doppler frequencies at the edge of the beam. The total beam extent is then

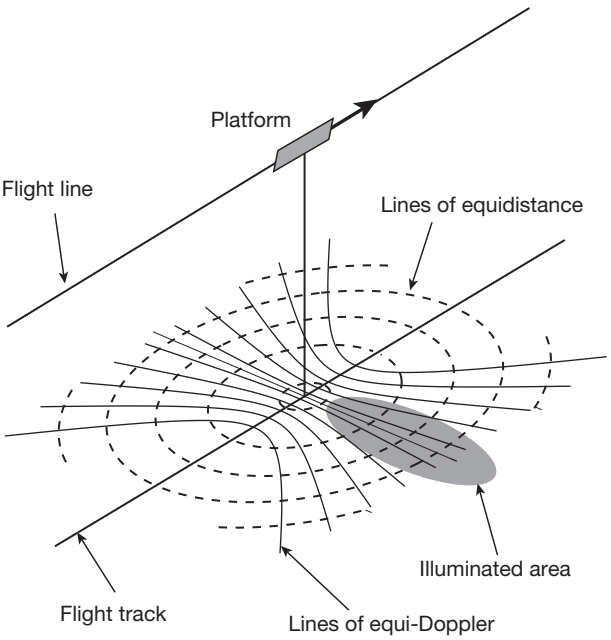


Figure 34 Range-Doppler coordinates of the surface as seen by a sensor at an instant of time. Targets in a band of Doppler frequencies contribute to the signal return. It is up to the azimuth matched-filtering operation to sort out the targets into unique Doppler bins.

$$2\theta_{\text{b.s.}} = X_{\text{ill}}/R = \lambda R/(LR) = \lambda/L \qquad [77]$$

as expected. Thus, the Doppler bandwidth is $f_{\text{D,ill}}$ and by this derivation is also $2v(2\sin\theta_{\text{b.s.}})/\lambda = 2v/L$.

At any given instant, the received signal is composed of the full complement of Doppler frequencies because targets are distributed throughout the beam, each with their own Doppler frequency. In forming the synthetic aperture at some reference time, we are rearranging each of the targets in the range-Doppler image to lie at the Doppler frequency appropriate to that geometry at that reference time. This is not immediately obvious from the signal-theoretical approach to azimuth filtering that was just derived. To be more explicit, we pick a collection of pulses that will be used to form our synthetic aperture, noting that this establishes a position in space at a time instant as a reference for defining Doppler frequencies. We can also model the ground as a collection of point targets arranged in azimuth, such that

$$r_{\text{zc}\delta,\text{tot}} = \sum_{k=1}^{N} A_k e^{-j\pi(x-x_k)^2/F^2} \operatorname{sinc}\frac{\pi}{\Delta R}\left(R - R_0 - (x-x_k)^2/2R_0\right) \qquad [78]$$

For the sake of this discussion, we allow the sinc function to become a delta function in range and ignore its variability with x, which is small and correctable. Thus,

$$r_{\text{zc}\delta} = \sum_{k=1}^{N} A_k e^{-j\pi(x-x_k)^2/F^2} \qquad [79]$$

The azimuth compression process can be written then as

$$\begin{aligned} r_{\text{zcc}\delta}(x) &= \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} r_{\text{zc}\delta}(x') e^{j\pi(x'-x)^2/F^2}\,dx' \\ &= \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} \sum_{k=1}^{N} A_k e^{-j\pi(x'-x_k)^2/F^2} e^{j\pi(x'-x)^2/F^2}\,dx' \\ &= e^{j\pi x^2/F^2} \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} \sum_{k=1}^{N} A_k e^{-j\pi x_k^2/F^2} e^{j2\pi x' x_k/F^2} e^{-j2\pi x' x/F^2}\,dx' \\ &= \sum_{k=1}^{N} e^{j\pi x^2/F^2} \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} A'_k e^{-j2\pi x'\left(f_{\text{Dx}} - f_{k,\text{Dx}}\right)}\,dx' \end{aligned} \qquad [80]$$

where $f_{\text{Dx}}(x) = x/F^2$ is the Doppler spatial frequency. From this arrangement of terms, we can recognize that the compressed azimuth signal is just the sum of Fourier transforms of the individual complex exponentials of complex amplitude A'_k. Because the argument of the exponential is the Doppler frequency, we see the transform domain is the Doppler domain, and each exponential (being representative of a scatterer at a particular azimuth location x_k relative to the reference location) transforms to an individual Doppler frequency $f_{k'\text{Dx}}$. Thus, for a general continuous surface, the summation becomes an integration, and we see the compression process consists of focusing the array by applying a quadratic phase correction, followed by a transform that sorts the scatterers into correct Doppler bins for that time reference.

The reason we represent (and in fact execute) the imaging as a convolution in spatial variables is because the processed data the sample spacing independent of range on input and output, which is often desirable. In the focusing Fourier transform approach, the Doppler sample spacing is controlled by the implementation of Fourier transform, which is usually accomplished by fast Fourier transform (FFT). In this case, if the input data are sampled at a spacing $L/2$, the Doppler spacing will be $2/L/N$, where N is the number of points in the FFT.

A.1.5 Other Doppler considerations

If the antenna pattern was isotropic and we have a continuous time system, the phase and Doppler of a single point target would as depicted in Figure 35. Imposing an antenna pattern where the antenna pattern is pointing 'broadside,' that is, orthogonal to the velocity vector, the phase and Doppler history would be truncated according to the extent of the beam (see same figure). It is most commonly the case that the finite extent beam is 'squinted' forward or back with respect to the velocity vector of the sensor. In an aircraft, this is most commonly due to cross winds causing the fuselage to crab. In spacecraft, there is a natural latitude dependent angular offset between the velocity vector of the inertial orbit and the rotating Earth below causing a natural squint (Figure 36). The effect of this squint on the phase and Doppler history of a target is shown in Figure 37. Note the shape of the phase or Doppler does not change, just the range of phases or frequencies considered. Thus, there is a unique relationship between the pointing and the location of the scatterers. In processing data, the extent of the antenna beam (Doppler bandwidth) and the *centroid* of the Doppler, or the degree of squint as represented by the center of the Doppler bandwidth, are specified in one way or another. Figure 38 shows that the spectra are shifted to be centered on the Doppler spread associated with a particular

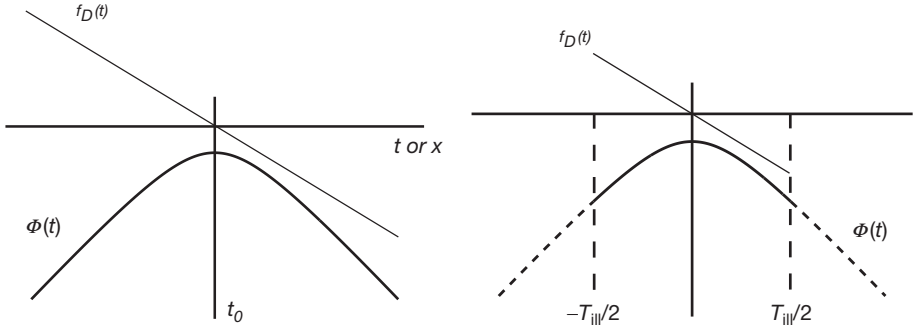


Figure 35 Phase and Doppler history over time for a point target. At left, the trace for a system with an isotropic antenna: the trace keeps going. At right, the finite illumination time of a finite dimension antenna truncates the phase and Doppler history.

squint. It is important to know the squint to ensure that the processing is centered on the proper part of the azimuth spectrum. The centroid specified in processing defines direction from which the ground is imaged after processing.

The squint can be derived from information provided by the spacecraft and radar manufacturer. Often, spacecrafts have sensitive position and attitude sensors that record the information needed to calculate the direction of the antenna boresight. When this information is not available, the Doppler centroid itself can be estimated from the data. This estimated centroid can be used directly in the processing. The centroid generally varies across range, so often the centroid is specified as a function of range.

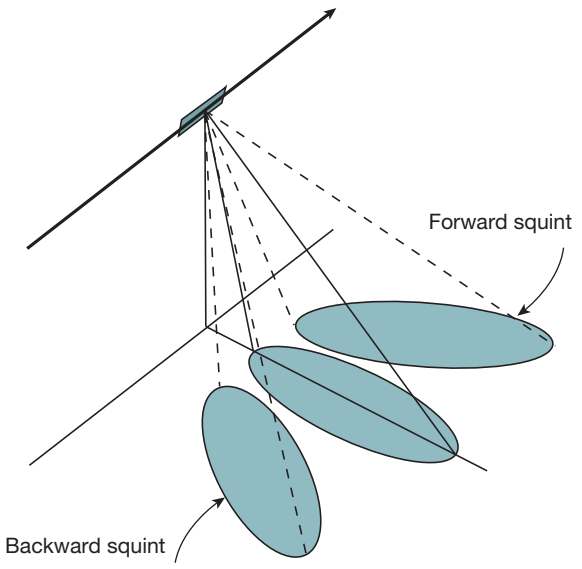


Figure 36 Illustration of forward and backward squint.

To understand the arrangement of spectral energy for a collection of scatterers, consider a collection of point target scatterers all at the same range, but at different azimuths:

$$\begin{aligned} r_{zc\delta,\mathrm{tot}} = {} & A_0 e^{-j\pi(x-x_0)^2/F^2} \operatorname{sinc}\frac{\pi}{\Delta R}\left(R - R_0 - (x-x_0)^2/2R_0\right) \\ & + A_1 e^{-j\pi(x-x_1)^2/F^2} \operatorname{sinc}\frac{\pi}{\Delta R}\left(R - R_0 - (x-x_1)^2/2R_0\right) \\ & + \cdots \\ & + A_n e^{-j\pi(x-x_n)^2/F^2} \operatorname{sinc}\frac{\pi}{\Delta R}\left(R - R_0 - (x-x_n)^2/2R_0\right) \end{aligned} \quad [81]$$

If we ignore the antenna pattern modulation of target energy and denote the azimuth spectrum of an individual scatter response as

$$F(\omega) = \mathrm{F.T.}\left\{ e^{-j\pi x^2/F^2} \operatorname{sinc}\frac{\pi}{\Delta R}\left(R - R_0 - x^2/2R_0\right)\right\} \quad [82]$$

where F.T. denotes the Fourier transform, the spectrum of $r_{zc\delta,\mathrm{tot}}$ would be

$$R_{zc\delta,\mathrm{tot}}(\omega) = F(\omega)\left(A_0 e^{-j\omega x_0} + A_1 e^{-j\omega x_1} + \cdots + A_n e^{-j\omega x_n}\right) \quad [83]$$

Thus, except for a phase ramp in azimuth frequency space, all scatterers have the same spectrum. Clearly, the process of applying the matched filter neutralizes this common component, and the linear phase terms in frequency distinguish the

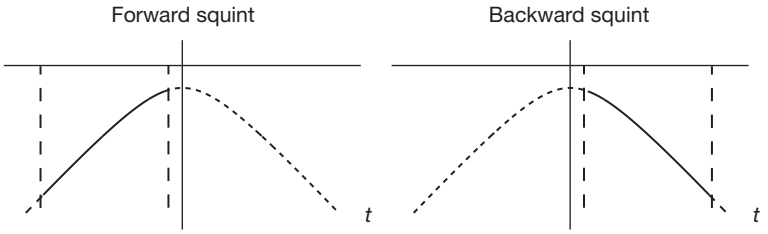


Figure 37 Phase history over time for a point target. At left, the trace for a forward squinted system. At right, the trace for a backward squinted system. Only phase is shown, but with reference to the previous figures, the Doppler history is similarly constrained. Forward (backward) squinted systems see more positive (negative) Doppler frequencies.

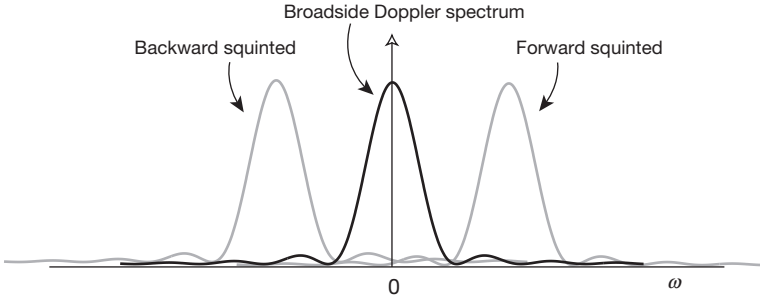


Figure 38 Hypothetical Doppler spectra of broadside, forward and backward squinted signals. For a collection of scatterers over the beam pointed with a particular squint, all scatterers experience the same Doppler frequency spread, so the spectra would be as shown, controlled in extent and magnitude by the antenna pattern in azimuth.

targets in position upon inverse transformation. This again shows the spatial-Doppler duality. The x dependence of the energy's localization in range $x^2/2R_0$ is called the range migration and is seen to be part of the function $F(\omega)$ in the frequency domain. Hence, any effects this may introduce in can be dealt with for all targets simultaneously by manipulating the spectrum.

We ignored the *range migration* term $(x-x_0)2/2R_0$ in the argument to the sinc function in range, but frequently this is not advisable. The processing of data on an orthogonal grid is enforced by the need for high-speed computations, and the most efficient processing methods are those that are separable in range and azimuth and for which FFT-based convolutions can be performed. For broadside imaging, the term can be ignored when $X_{\text{ill}}^2/2R_0$ is a fraction of the range resolution ΔR. In words, when the migration over the full extent of the azimuth matched filter (equivalent to the beam extent on the ground) is smaller than a range resolution element, the azimuth matched filter will catch all the energy for a given scatterer in a single range bin.

A.1.6 The sampled azimuth spectrum and range migration

Since the range migration correction terms depend on the Doppler frequency, one must know the correct Doppler to properly compress the imagery. This is also true for computing the correct azimuth matched filter; however, the sampled nature of the azimuth signal adds an interesting wrinkle to the processing that is often confusing. Since the azimuth signals in inherently derived from pulses acquired along-track, the azimuth signal is a discrete time signal sampled at the PRF. The actual discrete time spectrum is then a circular spectrum, with all energy at frequencies higher than the PRF aliased into the range of one PRF (Figure 39). Thus, without knowledge of the geometry of the spacecraft and the observation, it is often difficult to know the actual Doppler centroid, since one usually estimates it from the spectrum only in the range of the PRF. One can process the data with an incorrect Doppler centroid, off by some number of PRFs (also called *ambiguities*), for instance, and the imagery will look almost correct. Since the signal is sampled, the matched-filter spectrum for the actual centroid will be the same as for the incorrect centroid as long as the centroid is wrong by an integer number of ambiguities. However, the range migration correction is in terms of absolute Doppler, so if the wrong ambiguity is used, the images will be poorly focused in range (because energy is spread in range still for each scattered) and in azimuth (because the compression uses incomplete information for each scatterer).

If the attitude of the platform is known, then the Doppler can be computed from

$$f_{\text{D}} = -\frac{2\,\vec{v}\cdot\hat{l}}{\lambda} = -\frac{2v}{\lambda}\sin\theta\sin\phi_{\text{sq}} \qquad [84]$$

where $\vec{v}$ is the velocity vector, $\hat{l}$ is the look vector from antenna to target, θ is the look angle, and ϕ_{sq} is the squint angle. If the attitude parameters are not known, then one has to estimate the ambiguity. This can be done by trial and error – examine the focus for a variety of integer ambiguity choices – or by one of several automated techniques. A popular method of ambiguity estimation is to split the azimuth spectrum into two pieces and process each with a particular Doppler ambiguity assumption. If the correction is the right one, both side bands will be adjusted to the correct range and the two images produced will be properly registered in range. If the correction is not right, then there will be a residual shift between the two images in range. From the magnitude of this shift, it is often possible to estimate the correct ambiguity. Unfortunately, the imagery is lower resolution in azimuth and defocused in range, so estimating the shift is sometimes difficult.

A.2 The Derivation of the Interferogram from Two Images

As shown leading up to eqn [72], for a general complex scene $\Gamma(x, r)$, the SAR image after compression is given by convolution of r with the point target response

$$\Gamma_{\text{zcc}}(x', r') = \iint \Gamma(x, r) r_{\text{zcc}\delta}(x', r'; x, r)\,\text{d}x\,\text{d}r$$
$$= \int \text{d}x \int \text{d}r \Gamma(x, r) \text{e}^{-\text{j}4\pi r/\lambda} \cdot \text{sinc}\left(\frac{\pi}{\Delta R}(r' - r)\right) \text{sinc}\left(\frac{\pi}{\Delta X}(x' - x)\right) \qquad [85]$$

which can be verified by substituting $\Gamma(x, r) = \delta(x - x_0, r - R_0)$ to recover the impulse response $r_{\text{zcc}\delta}$.

The SAR image Γ_{zcc} is the convolution of the actual reflectivity with a two-dimensional function similar to a delta function but with finite width (the sinc function). This convolution smears out the intrinsic reflectivity as a point spread function in optics. If the system bandwidths were to become infinite, the sinc functions would become delta functions

$$\text{sinc}\left(\frac{\pi}{\Delta R}(r' - r)\right) \to \delta(r' - r) \quad \text{as } \Delta R \to 0 \qquad [86]$$

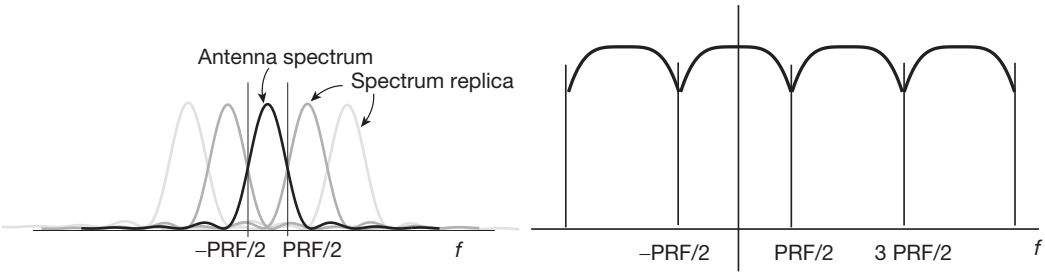


Figure 39 The azimuth signal is sampled at the PRF. The resulting discrete time signal in azimuth has a circular spectrum with all energy at frequency higher than the PRF aliased into the circular bandwidth of the PRF. Left illustrates the replicas of the continuous spectrum that are aliased, right the resulting circular spectrum.

$$\mathrm{sinc}\left(\frac{\pi}{\Delta X}(x'-x)\right) \to \delta(x'-x) \quad \text{as } \Delta X \to 0 \qquad [87]$$

such that

$$\Gamma_{\mathrm{zcc}}(x', r') \to \Gamma(x', r')\mathrm{e}^{-\mathrm{j}4\pi r'/\lambda} \qquad [88]$$

Now consider two observations of the reflectivity acquired from slightly different ranges. For the original observation at r

$$\Gamma_{\mathrm{zcc},1}(x', r') = \int \mathrm{d}x \int \mathrm{d}r \Gamma(x, r)\mathrm{e}^{-\mathrm{j}4\pi r/\lambda} \cdot \mathrm{sinc}\left(\frac{\pi}{\Delta R}(r'-r)\right) \mathrm{sinc}\left(\frac{\pi}{\Delta X}(x'-x)\right) \qquad [89]$$

For an observation at $r+\delta r$, the point target response is now shifted, but the scene must still be referenced to the original range r:

$$r_{\mathrm{zcc}\delta}(x', r'; x, r+\delta r) = \mathrm{e}^{-\mathrm{j}4\pi(r+\delta r)/\lambda} \cdot \mathrm{sinc}\left(\frac{\pi}{\Delta R}(r'-(r+\delta r))\right) \mathrm{sinc}\left(\frac{\pi}{\Delta X}(x'-x_0)\right) \qquad [90]$$

$$\Gamma_{\mathrm{zcc},2}(x', r') = \int \mathrm{d}x \int \mathrm{d}r \Gamma(x, r)\mathrm{e}^{-\mathrm{j}4\pi(r+\delta r)/\lambda} \cdot \mathrm{sinc}\left(\frac{\pi}{\Delta R}(r'-(r+\delta r))\right) \mathrm{sinc}\left(\frac{\pi}{\Delta X}(x'-x)\right) \qquad [91]$$

Again letting the bandwidths tend to infinity, we get

$$\Gamma_{\mathrm{zcc},1}(x', r') \to \Gamma(x', r')\mathrm{e}^{-\mathrm{j}4\pi r'/\lambda} \qquad [92]$$

$$\Gamma_{\mathrm{zcc},2}(x', r') \to \Gamma(x', r'-\delta r)\mathrm{e}^{-\mathrm{j}4\pi r'/\lambda} \qquad [93]$$

Note a shift in the registration of the true scene reflectivity.

$$\Gamma_{\mathrm{zcc},2}(x', r'+\delta r) \to \Gamma(x', r')\mathrm{e}^{-\mathrm{j}4\pi(r'+\delta r)/\lambda} \qquad [94]$$

When the true scene reflectivity is aligned, there is a phase difference between the reconstructed phase proportional to δr, a geometric term. The true scene reflectivity is a complex number with random phase. However, after we have aligned the scene reflectivity by shifting $\Gamma_{\mathrm{zcc},2}(x', r')$ by δr, or equivalently by looking up the value of $\Gamma_{\mathrm{zcc},2}$ at location $r'+\delta r$, then the scene phase is common to both observations. We can form the product

$$I(x', r') = \Gamma_{\mathrm{zcc},1}(x', r')\Gamma^*_{\mathrm{zcc},2}(x', r'+\delta r) = |\Gamma(x', r')|^2 \mathrm{e}^{\mathrm{j}4\pi(\delta r)/\lambda} \qquad [95]$$

The function $I(x', r')$ is the *interferogram*, a complex quantity, the phase of which is just a geometric term related to the range difference δr (when bandwidths are infinite) between the two images. The range difference can be caused by a vantage point difference as described here and illustrated at the left panel in Figure 3 or by a shift in the scene location (at right in the figure) or a combination of the two. If the scene shifts rather than the vantage point, then the sign of the range difference changes, but the form of the interferogram is the same.

A.3 Speckle

Radar images are naturally noise because of the coherent interaction of the electromagnetic wave with scatterers within a resolution cell known as speckle. Speckle is an important part of the radar literature, as it is an important limitation of the imagery that many people have worked hard to minimize. It is also an important concept for interferometry, which exploits the spatially fixed properties of speckle to extract phase differences between images. The normalized backscatter cross section σ_0 is an average quantity for a naturally varying random surface. For homogenous surfaces made up of a continuum of scatterers, the backscatter cross section is a measure of the roughness of the surface and its natural reflectivity, as shown in a previous section. Viewing the SAR imaging process as a lens (Figure 40), after imaging, the random (complex) surface reflectivity is convolved with the sinc function impulse response of the imaging system accounting for the random differential path length introduced by surface roughness.

In a given resolution element, we consider the sum of the collection of small independent scatterers, where each scatterer has its own amplitude and phase

$$a\mathrm{e}^{\mathrm{j}\theta} = \sum_{k=1}^{N} \alpha_k \mathrm{e}^{\mathrm{j}\phi_k} \qquad (96)$$

If α_k and ϕ_k are statistically independent of each other and α_k are identically distributed with mean $\overline{\alpha}$ and second moment $\overline{\alpha^2}$ and ϕ_k are uniformly distributed in the interval $[-\pi, \pi]$, then one can show that for large N,

$$r = a\cos\theta, \quad i = a = \sin\theta \qquad [97]$$

are Gaussian-distributed zero-mean random variables with joint probability density function

$$p_{\mathrm{RI}}(r, i) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{-\left(r^2+i^2\right)/2\sigma^2} \qquad [98]$$

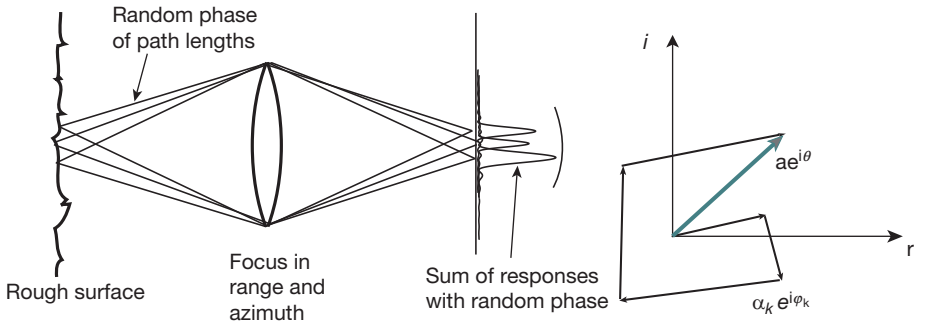


Figure 40 Illustration of the concept of radar imaging of a random surface (left) and the coherent sum of independent scatterers as a phasor (right).

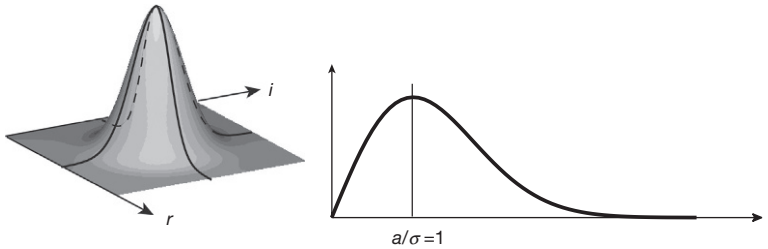


Figure 41 The real and imaginary parts of a SAR image are circularly Gaussian (left). The amplitude of the SAR pixel is Rayleigh distributed, with peak at $a=\sigma$.

where $\sigma^2=\overline{\alpha^2}/2$ (Figure 41). The variances of the real and imaginary parts of the SAR image are related to the mean power of the intrinsic scatterers but are zero-mean random variables themselves. As such for homogeneous areas, averaging many complex pixels together in a region will eventually reduce the signal to zero on average, and not improve the quality of the image. This is a common mistake among beginners. It is the amplitude of the image, a, or its intensity (power), $I=a^2$, that interests us.

For the amplitude $a=\sqrt{r^2+i^2}$, we can use the rules of transformation of random variables (multiply the distribution by the Jacobian of the transformation and substitute in the new variables) to get the distribution

$$p_A(a)=\frac{a}{\sigma^2}e^{-a^2/2\sigma^2},\quad a\geq 0 \qquad [99]$$

which is known as the Rayleigh distribution and is illustrated in Figure 41. The phase is uniformly distributed over the interval $(-\pi,\pi)$

$$p_\Theta(\theta)=\frac{1}{2\pi},\quad -\pi<\theta<\pi \qquad [100]$$

The mean and variance of the amplitude are

$$\overline{a}=\sqrt{\frac{\pi}{2}}\sigma,\quad \sigma_a^2=\left(2-\frac{\pi}{2}\right)\sigma^2 \qquad [101]$$

Clearly the mean amplitude of the image scales with the variability in the image, which is related to the mean backscattered power $\overline{\alpha^2}$ of the scatterers comprising the resolution element. The brighter the image, the noisier the image, and vice versa. Thus, at some signal level such that the signal level returned is sufficiently larger than the thermal noise in an image, having a more powerful radar does not necessarily improve image quality.

The intensity, or power, of the image $I=a^2$ has an exponential distribution

$$p_I(I)=\frac{1}{\overline{I}}e^{-I/\overline{I}},\quad I\geq 0 \qquad [102]$$

with mean and variance

$$\overline{I}=2\sigma^2,\quad \sigma_I^2=4\sigma^4 \qquad [103]$$

The mean power is equal to its standard deviation, as expected from the mean amplitude scaling with σ. In the display and analysis of an image with only 'one look,' meaning processed at full resolution using the full synthetic aperture, the SNR of the intensity image is just $1\left(\overline{I}^2/\sigma_I^2=1\right)$, the brighter the image, the noisier. In the design of a radar system, one must have sufficient power to receive a return from the surface above the thermal noise (satisfying the radar equation where $\overline{\alpha^2}$ is related to the backscattering cross section), but not so high that we are wasting power. For interferometry, we will soon see however that more power in general is better!

The radiometric resolution is defined as the ability to discriminate surfaces of different brightnesses. It has been defined to be

$$Q=10\log\left(\frac{\overline{I}+\sigma_I}{\overline{I}-\sigma_I}\right) \qquad [104]$$

which is something like the distance apart in amplitude of discriminable amplitudes. For an SLC image, that distance is infinite. In order to improve this situation, it is often necessary to average many pixels together. The act of averaging N independent (Gaussian, but also approximately true for non-Gaussian) random variables with identical distributions reduces the variance by a factor of N. Thus, for

$$I'=\frac{1}{N}\sum_{k=1}^{N}I_k \qquad [105]$$

we have $\overline{I'}=\overline{I}$, $\sigma_{I'}^2=\sigma_I^2/N$, and the radiometric resolution becomes

$$Q=10\log\left(\frac{\overline{I'}+\sigma_{I'}}{\overline{I'}-\sigma_{I'}}\right)=10\log\left(\frac{\sqrt{N}+1}{\sqrt{N}-1}\right) \qquad [106]$$

The resolution distance approaches 0 for arbitrarily large N. This averaging technique is often called 'taking looks' because it can be shown to be equivalent to segmenting the azimuth spectrum into separate frequency bands, forming a power image from each band, and then stacking the images in an average. Since azimuth band corresponds to a particular look direction relative to a target on the ground, this technique is called taking looks.

Of course, averaging together a large number of pixels will lower the intrinsic resolution of the image, so designing the system to have a larger bandwidth is desirable if radiometric resolution is important. Note also that there is a limit to how far one can go to improve the statistics. The improvement in the preceding text assumes that all the pixels come from the same distribution. If the scene is variable in intrinsic brightness, not all the benefits will be realized.

If there is a bright target within a resolution cell, then there is a mean value associated with the Gaussian random variables (Figure 42). This complex random variable has a Rician distribution, given by

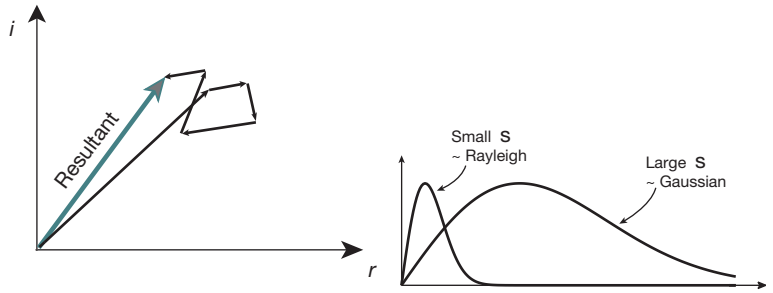


Figure 42 A phasor diagram with a bright specular target (within a scattering cell) (left). The Rician distribution of the amplitude (right).

$$p_{\mathrm{A}}(a)=\frac{a}{\sigma^2}\mathrm{e}^{-(a^2+s^2)/2\sigma^2}I_0\left(\frac{as}{\sigma^2}\right),\quad a\geq 0 \qquad [107]$$

where I_0 is the modified Bessel function of the first kind and s is the specular target strength. This distribution approaches a Gaussian distribution for larges relative to σ. The distribution of the phase, which is uniform for $s=0$, becomes more peaked as s grows, eventually approximating a Gaussian also,

$$p_{\Theta}(\theta)\rightarrow\frac{s/\sigma}{\sqrt{2\pi}}\mathrm{e}^{-s^2\theta^2/2\sigma^2} \qquad [108]$$

Often, specular targets are used for the calibration of the radar. If the specular target is not sufficiently bright, then the amplitude and phase will be corrupted by the background random scatterers.

Another way to model the surface is in terms of a general reflectivity function that is statistical in nature. Since

$$\Gamma_{\mathrm{zcc}}(x',r')=\iint \mathrm{d}r\,\mathrm{d}x\,\Gamma(x,r)\mathrm{sinc}\left(\frac{\pi}{\Delta X}(x'-x)\right)\mathrm{sinc}\left(\frac{\pi}{\Delta R}(r'-r)\right) \qquad [109]$$

if Γ is a statistical function, the power that is meaningful is the expected power scattered back from the resolution cell, which is related by scale factors and geometric normalizations to $E\{\Gamma\Gamma^*\}$. Writing this out explicitly, we have

$$\begin{aligned}E\left\{\Gamma_{\mathrm{zcc}}\Gamma^*_{\mathrm{zcc}}\right\}(x',r')=&\int \mathrm{d}x_1\int \mathrm{d}x_2\int \mathrm{d}r_1\int \mathrm{d}r_2\,E\{\Gamma(x_1,r_1)\Gamma^*(x_2,r_2)\}\\&\cdot\mathrm{sinc}\left(\frac{\pi}{\Delta X}(x'-x_1)\right)\mathrm{sinc}\left(\frac{\pi}{\Delta X}(x'-x_2)\right)\\&\cdot\mathrm{sinc}\left(\frac{\pi}{\Delta R}(r'-r_1)\right)\mathrm{sinc}\left(\frac{\pi}{\Delta R}(r'-r_2)\right)\end{aligned} \qquad [110]$$

If the scatterers that comprise the reflectivity are uncorrelated, then the expected value of the product of the surface with itself when not coaligned would be zero:

$$E\{\Gamma(x_1,r_1)\Gamma^*(x_2,r_2)\}=E\{\Gamma(x_1,r_1)\}E\{\Gamma^*(x_2,r_2)\}=0 \qquad [111]$$

However, when the surface is self-aligned,

$$E\{\Gamma(x_1,r_1)\Gamma^*(x_1,r_1)\}=|\Gamma(x_1,r_1)|^2 \qquad [112]$$

More succinctly,

$$E\{\Gamma(x_1,r_1)\Gamma^*(x_2,r_2)\}=E\left\{|\Gamma(x_1,r_1)|^2\right\}\delta(x_1-x_2)\delta(r_1-r_2) \qquad [113]$$

Using this expression in eqn [110] simplifies the integrals

$$\begin{aligned}E\left\{\Gamma_{\mathrm{zcc}}\Gamma^*_{\mathrm{zcc}}\right\}(x',r')=&\int \mathrm{d}x\int \mathrm{d}r E\left\{|\Gamma(x,r)|^2\right\}\cdot\mathrm{sinc}^2\left(\frac{\pi}{\Delta X}(x'-x)\right)\\&\mathrm{sinc}^2\left(\frac{\pi}{\Delta R}(r'-r)\right)=E\left\{|\Gamma|^2\right\}\Delta X\Delta R\end{aligned} \qquad [114]$$

For homogeneous surfaces, we can expect the mean power to be essentially constant over the resolution cell, so we can take the expectation outside the integration. Also the variables x' and r' can profitably be set to zero since there is no dependence on space. Then the integration over the sinc functions produces a constant, which is proportional to the area under them. This turns out to be $\triangle X$ and $\triangle R$. So we see that $E\{\Gamma_{\mathrm{zcc}}\Gamma^*_{\mathrm{zcc}}\}$ is related to the normalized backscattering cross section σ_0 through the function Γ.

A.4 Doppler and the Interferometric Baseline

The precise definition of interferometric baseline and phase, and consequently the topographic mapping process, depends on how the SAR data comprising the interferometer are processed. Consequently, a brief overview of the salient aspects of SAR processing is in order.

Processed data from SAR systems are sampled images. Each sample, or pixel, represents some aspect of the physical process of radar backscatter. A resolution element of the imagery is defined by the spectral content of the SAR system. Fine resolution in the range direction is achieved typically by transmitting pulses of either short time duration with high peak power or a longer time duration with a wide, coded signal bandwidth at lower peak transmit power. Resolution in range is inversely proportional to this bandwidth. In both cases, the received echo for each pulse is sampled at the required radar signal bandwidth.

For ultranarrow pulsing schemes, the pulse width is chosen at the desired range resolution, and no further data manipulation is required. For coded pulses, the received echoes are typically processed with a matched-filter technique to achieve the desired range resolution. Most spaceborne platforms use chirp encoding to attain the desired bandwidth and consequent range resolution, where the frequency is linearly changed across the pulse.

Resolution in the azimuth, or along-track, direction, parallel to the direction of motion, is achieved by synthesizing a large antenna from the echoes received from the sequence of pulses illuminating a target. The pulses in the synthetic aperture contain an unfocussed record of the target's amplitude and phase history. To focus the image in azimuth, a digital 'lens' that mimics the imaging process is constructed and is applied

by matched filtering. Azimuth resolution is limited by the size of the synthetic aperture, which is governed by the amount of time a target remains in the radar beam. The azimuth beamwidth of an antenna is given by $\theta_{bw} \equiv k\lambda/L$, where λ is the wavelength, L is the antenna length, and k is a constant that depends on the antenna ($k = 1$ is assumed in this chapter). The size of the antenna footprint on the ground in the azimuth direction is approximately given by

$$l_{az} = \rho\theta_{bw} = \rho\frac{\lambda}{L} \qquad [115]$$

where ρ is the range to a point in the footprint.

During the time a target is in the beam, the range and angular direction to the target are changing from pulse to pulse, as shown in Figure 31. To generate a SAR image, a unique range or angle must be selected from the family of ranges and angles to use as a reference for focusing the image. Once selected, the target's azimuth and range position in the processed image is uniquely established. Specifying an angle for processing is equivalent to choosing a reference Doppler frequency. The bold dashed line from pulse $N-2$ to the target in Figure 31 indicates the desired angle or Doppler frequency at which the target will be imaged. This selection implicitly specifies the time of imaging and therefore the location of the radar antenna. This is an important and often ignored consideration in defining the interferometric baseline. The baseline is the vector connecting the locations of the radar antennas forming the interferometer; since these locations depend on the choice of processing parameters, so does the baseline. For two-aperture cross track interferometers, this is a subtle point; however, for repeat-track geometries where the antenna pointing can be different from track to track, careful attention to the baseline model is essential for accurate mapping performance.

Acknowledgments

Radar images used in this chapter were provided by ESA under multiple proposals (M.S. PI) as well as via the NSF/NASA/USGS-sponsored WInSAR data consortium under the auspices of UNAVCO. The authors thank Eric Fielding and Paul Lundgren for helpful comments on this manuscript. Mark Simons thanks Rowena Lohman, Matthew Pritchard, Yuri Fialko, Eric Hetland, Pablo Muse, Piyush Agram, and Romain Jolivet for their extensive discussions on the optimal use of InSAR data. Paul A. Rosen thanks Gilles Peltzer for several of the graphics included in this chapter that they jointly developed previously for teaching purposes (Contribution number 9166, Geological and Planetary Sciences, California Institute of Technology, Pasadena CA 91125). This research was supported in part by the Gordon and Betty Moore Foundation. This is Caltech Tectonic Observatory Contribution 67. For P.A.R., this chapter was prepared at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not constitute or imply its endorsement by the US government or the Jet Propulsion Laboratory, California Institute of Technology.

References

Agram PS, Jolivet R, Riel B, et al. (2013) New radar interferometric time series analysis toolbox released. *Eos, Transactions of the American Geophysical Union* 94: 69–70.

Alsdorf DE, Smith LC, and Melack JM (2001) Amazon floodplain water level changes measured with interferometric SIR-C radar. *IEEE Transactions on Geoscience and Remote Sensing* 39: 423–431.

Argus DF, Heflin MB, and Peltzer G (2005) Interseismic strain accumulation and anthropogenic motion in metropolitan Los Angeles. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2003JB002934.

Ballatore P (2006) Synthetic aperture radar interferometry: Separation of atmospheric artifacts from effects due to the topography and the terrain displacements. *Earth, Planets and Space* 58: 927–935.

Bamler R and Eineder M (1996) ScanSAR processing using standard high-precision SAR algorithms. *IEEE Transactions on Geoscience and Remote Sensing* 34: 212–218.

Bamler R and Hartl P (1998) Synthetic aperture radar interferometry. *Inverse Problems* 14: 1–54.

Bawden GW, Thatcher W, Stein RS, Hudnut KW, and Peltzer G (2001) Tectonic contraction across Los Angeles after removal of groundwater pumping effects. *Nature* 412: 812–815.

Bechor NBD and Zebker HA (2006) Measuring two-dimensional movements using a single InSAR pair. *Geophysical Research Letters* 33. http://dx.doi.org/10.1029/2006GL026883.

Bell JW, Amelung F, Ramelli AR, and Blewitt G (2002) Land subsidence in Las Vegas, Nevada, 1935–2000: New geodetic data show evolution, revised spatial patterns, and reduced rates. *Environmental and Engineering Geoscience* 8: 155–174.

Berardino P, Fornaro G, Lanari R, and Sansosti E (2002) A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing* 40: 2375–2383.

Biggs J, Wright T, Lu Z, and Parsons B (2007) Multi-interferogram method for measuring interseismic deformation: Denali Fault, Alaska. *Geophysical Journal International* 170: 1165–1179.

Born M and Wolf E (1989) *Principles of Optics*, 6th edn. Oxford: Pergamon Press.

Bürgmann R, Hilley G, Ferretti A, and Novali F (2006) Resolving vertical tectonics in the San Francisco Bay Area from permanent scatterer InSAR and GPS analysis. *Geology* 34: 221–224.

Bürgmann R, Rosen PA, and Fielding EJ (2000) Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. *Annual Review of Earth and Planetary Sciences* 28: 169–209.

Cavalié O, Doin MP, Lasserre C, and Briole P (2007) Ground motion measurement in the Lake Mead area, Nevada, by differential synthetic aperture radar interferometry time series analysis: Probing the lithosphere rheological structure. *Journal of Geophysical Research* 112: B03403.

Cavalié O, Lasserre C, Doin MP, et al. (2008) Measurement of interseismic strain across the Haiyuan fault (Gansu, China), by InSAR. *Earth and Planetary Science Letters* 275: 246–257.

Chen CW and Zebker HA (2001) Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization. *Journal of the Optical Society of America A* 18: 338–351.

Cloude SR (2006) Polarization coherence tomography. *Radio Science* 1944-799X. 41: RS4017. http://dx.doi.org/10.1029/2005RS003436.

Cloude SR and Papathanassiou KP (1998) Polarimetric SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 36: 1551–1565.

Colesanti C, Ferretti A, Novali F, Prati CC, and Rocca F (2003) SAR monitoring of progressive and seasonal ground deformation using the permanent scatterers technique. *IEEE Transactions on Geoscience and Remote Sensing* 41: 1685–1701.

Davidson GW and Bamler R (1999) Multiresolution phase unwrapping for SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 37: 163–174.

De Zan F and Monti Guarnieri AM (2006) TOPSAR: Terrain observation by progressive scans. *IEEE Transactions on Geoscience and Remote Sensing* 44: 2352–2360.

Dee DP, Uppala SM, Simmons AJ, et al. (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137: 553–597.

Delacourt C, Briole P, and Achache J (1998) Tropospheric corrections of SAR interferograms with strong topography: Application to Etna. *Geophysical Research Letters* 25: 2849–2852.

Dicaprio CJ and Simons M (2008) Importance of ocean tidal load corrections for differential InSAR. *Geophysical Research Letters* 35: 1–5.

Doin MP, Lasserre C, Peltzer G, Cavalié O, and Doubre C (2009) Corrections of stratified tropospheric delays in SAR interferometry: Validation with global atmospheric models. *Journal of Applied Geophysics* 69: 35–50.

Dong D, Fang P, Bock Y, Cheng MK, and Miyazaki S (2002) Anatomy of apparent seasonal variations from GPS-derived site position time series. *Journal of Geophysical Research* 107: 2075.

Dong D, Fang P, Bock Y, et al. (2006) Spatiotemporal filtering using principal component analysis and Karhunen–Loeve expansion approaches for regional GPS network analysis. *Journal of Geophysical Research* 111: B03405. http://dx.doi.org/10.1029/2005JB003806.

Dong D, Herring TA, and King RW (1998) Estimating regional deformation from a combination of space and terrestrial geodetic data. *Journal of Geodesy* 72: 200–214. http://dx.doi.org/10.1007/s001900050161.

Dong D, Yunck T, and Heflin M (2003) Origin of the international terrestrial reference frame. *Journal of Geophysical Research* 108: 2200.

Doubre C and Peltzer G (2007) Fluid-controlled faulting process in the Asal Rift, Djibouti, from 8 yr of radar interferometry observations. *Geology* 35: 69–72.

Elachi C (1988) *Spaceborne Radar Remote Sensing: Applications and Techniques*. New York: IEEE Press.

Elliott JR, Biggs J, Parsons B, and Wright TJ (2008) InSAR slip rate determination on the Altyn Tagh Fault, northern Tibet, in the presence of topographically correlated atmospheric delays. *Geophysical Research Letters* 35: 1–5.

Emardson TR, Simons M, and Webb FH (2003) Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB001781.

Farr TG, Rosen PA, Caro E, et al. (2007) The shuttle radar topography mission. *Reviews of Geophysics* 45.

Feigl KL and Thurber CH (2009) A method for modelling radar interferograms without phase unwrapping: Application to the M 5 Fawnskin, California earthquake of 1992 December 4. *Geophysical Journal International* 176: 491–504.

Ferretti A, Fumagalli A, Novali F, Prati C, Rocca F, and Rucci A (2011) A new algorithm for processing interferometric data-stacks: Squeesar. *IEEE Transactions on Geoscience and Remote Sensing* 0196-2892. 49(9): 3460–3470. http://dx.doi.org/10.1109/TGRS.2011.2124465.

Ferretti A, Prati C, and Rocca F (2000) Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 38: 2202–2212.

Ferretti A, Prati C, and Rocca F (2001) Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 39: 8–20.

Fialko Y, Sandwell D, Simons M, and Rosen P (2005) Three-dimensional deformation caused by the Bam, Iran, earthquake and the origin of shallow slip deficit. *Nature* 435: 295–299.

Fialko Y, Simons M, and Agnew D (2001a) The complete (3-D) surface displacement field in the epicentral area of the 1999 m_w 7.1 Hector Mine earthquake, California, from space geodetic observations. *Geophysical Research Letters* 28: 3063–3066.

Fialko Y, Simons M, and Khazan Y (2001b) Finite source modelling of magmatic unrest in Socorro, New Mexico, and Long Valley, California. *Geophysical Journal International* 146: 191–200.

Fielding EJ, Talebian M, Rosen PA, et al. (2005) Surface ruptures and building damage of the 2003 Bam, Iran, earthquake mapped by satellite synthetic aperture radar interferometric correlation. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2004JB003299.

Fornaro G, Monti Guarnieri A, Pauciullo A, and De-Zan F (2006) Maximum likelihood multi-baseline SAR interferometry. *IEE Proceedings – Radar, Sonar and Navigation* 153: 279–288.

Foster J, Brooks B, Cherubini T, Shacat C, Businger S, and Werner CL (2006) Mitigating atmospheric noise for InSAR using a high resolution weather model. *Geophysical Research Letters* 33. http://dx.doi.org/10.1029/2006GL026781.

Fujiwara S, Rosen PA, Tobita M, and Murakami M (1998) Crustal deformation measurements using repeat-pass JERS-1 synthetic aperture radar interferometry near the Izu Peninsula, Japan. *Journal of Geophysical Research* 103: 2411–2426.

Gabriel AK, Goldstein RM, and Zebker HA (1989) Mapping small elevation changes over large areas: Differential radar interferometry. *Journal of Geophysical Research* 94: 9183–9191.

Gatelli F, Guarnieri AM, Parizzi F, Pasquali P, Prati C, and Rocca F (1994) The wave-number shift in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 32: 855–865.

Goldstein RM (1995) Atmospheric limitations to repeat-track radar interferometry. *Geophysical Research Letters* 22: 2517–2520.

Goldstein RM, Engelhardt H, Kamb B, and Frolich RM (1993) Satellite radar interferometry for monitoring ice sheet motion: Application to an Antarctic ice stream. *Science* 262: 1525–1530.

Goldstein RM and Zebker HA (1987) Interferometric radar measurement of ocean surface currents. *Nature* 328: 707–709.

Goldstein RM, Zebker HA, and Werner CL (1988) Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Science* 23: 713–720.

Goodman JW (1985) *Statistical Optics*. New York: Wiley-Interscience.

Gourmelen N and Amelung F (2005) Post-seismic mantle relaxation in the Central Nevada Seismic Belt. *Science* 310: 1473–1476.

Gourmelen N, Amelung F, and Lanari R (2010) Interferometric synthetic aperture radar–GPS integration: Interseismic strain accumulation across the Hunter Mountain fault in the eastern California shear zone. *Journal of Geophysical Research* 115: B09408. http://dx.doi.org/10.1029/2009JB007064.

Grandin R, Socquet A, Doin MP, Jacques E, de Chabalier JB, and King GCP (2010) Transient rift opening in response to multiple dike injections in the Manda Hararo rift (Afar, Ethiopia) imaged by time-dependent elastic inversion of interferometric synthetic aperture radar data. *Journal of Geophysical Research* 115: B09403. http://dx.doi.org/10.1029/2009JB006883.

Grell GA, Dudhia J, and Stauffer PJ (1995) A description of the fifth generation Penn State/NCAR Mesoscale Model (MM/5). NCAR Technical Note, National Center of the Atmospheric Research.

Guarnieri AM and Prati C (1996) ScanSAR focusing and interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 3: 1029–1038.

Hanssen RF (2001) *Radar Interferometry: Data Interpretation and Error Analysis*. Dordrecht, Netherlands: Kluwer.

Hetland EA, Muse P, Simons M, Lin YN, Agram PS, and DiCaprio CJ (2012) Multiscale InSAR Time Series (MInTS) analysis of surface deformation. *Journal of Geophysical Research* 117: B02404.

Holzner J and Bamler R (2002) Burst-mode and ScanSAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 40: 1917–1934.

Hooper A, Bekaert D, Spaans K, and Arikan M (2012) Recent advances in SAR interferometry time series analysis for measuring crustal deformation. *Tectonophysics* 0040-1951. 514–517: 1–13. http://dx.doi.org/10.1016/j.tecto.2011.10.013.

Hooper A, Zebker H, Segall P, and Kampes B (2004) A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. *Geophysical Research Letters* 31: L23611.

Jolivet R, Grandin R, Lasserre C, Doin MP, and Peltzer G (2011) Systematic InSAR tropospheric phase delay corrections from global meteorological reanalysis data. *Geophysical Research Letters* 38. http://dx.doi.org/10.1029/2011GL048757.

Jonsson S, Zebker H, Segall P, and Amelung F (2002) Fault slip distribution of the 1999 m_w 7.1 Hector Mine, California, earthquake, estimated from satellite radar and GPS measurements. *Bulletin of the Seismological Society of America* 92: 1377–1389.

Joughin I, Winebrenner DP, and Percival DB (1994) Probability density functions for multi-look polarimetric signatures. *IEEE Transactions on Geoscience and Remote Sensing* 32: 562–574.

Jung H, Won J, and Kim S (2009) An improvement of the performance of multiple-aperture SAR interferometry (MAI). *IEEE Transactions on Geoscience and Remote Sensing* 47: 2859–2869.

Klinger Y, Xu XW, Tapponnier P, Van der Woerd J, Lasserre C, and King G (2005) High-resolution satellite imagery mapping of the surface rupture and slip distribution of the M_W 7.8, 14 November 2001 Kokoxili earthquake, Kunlun Fault, northern Tibet, China. *Bulletin of the Seismological Society of America* 95: 1970–1987.

Lanari R, Hensley S, and Rosen PA (1998) Chirp-Z transform based SPECAN approach for phase-preserving ScanSAR image generation. *IEE Proceedings – Radar, Sonar and Navigation* 145: 254–261.

Lanari R, Lundgren P, Manzo M, and Casu F (2004a) Satellite radar interferometry time series analysis of surface deformation for Los Angeles, California. *Geophysical Research Letters* 31: L23613. http://dx.doi.org/10.1029/2004GL021294.

Lanari R, Mora O, Manunta M, Mallorqui JJ, Berardino P, and Sansosti E (2004b) A small baseline approach for investigating deformations on full resolution differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing* 42: 1377–1386.

Lee J-S, Hoppel KW, Mango SA, and Miller AR (1992) Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* 32: 1017–1028.

Levy F, Hsu Y, Simons M, LePrince S, and Avouac JP (2005) Distribution of coseismic slip for the 1999 Chi Chi Taiwan earthquake: New data and implications of varying 3D fault geometry. *Eos, Transactions of the AGU, Fall Meeting Supplement* 86: B1305.

Li ZW, Ding XL, and Liu G (2004) Modeling atmospheric effects on InSAR with meteorological and continuous GPS observations: Algorithms and some test results. *Journal of Atmospheric and Solar-Terrestrial Physics* 66: 907–917.

Li ZH, Fielding EJ, Cross P, and Muller JP (2006a) Interferometric synthetic aperture radar atmospheric correction: GPS topography-dependent turbulence model. *Journal of Geophysical Research* 111: B02404.

Li F and Goldstein RM (1990) Studies of multibaseline spaceborne interferometric synthetic aperture radars. *IEEE Transactions on Geoscience and Remote Sensing* 28: 88–97.

Li Z, Muller JP, Cross P, Albert P, Fischer J, and Bennartz R (2006b) Assessment of the potential of MERIS near-infrared water vapour products to correct ASAR interferometric measurements. *International Journal of Remote Sensing* 27: 349–365.

Li ZH, Muller JP, Cross P, and Fielding EJ (2005) Interferometric synthetic aperture radar (InSAR) atmospheric correction: GPS, moderate resolution imaging spectroradiometer (MODIS), and InSAR integration. *Journal of Geophysical Research* 110: B03410.

Li ZW, Xu WB, Feng GC, et al. (2012) Correcting atmospheric effects on InSAR with MERIS water vapour data and elevation-dependent interpolation model. *Geophysical Journal International* 189: 898–910.

Lin YN, Simons M, Hetland EA, Muse P, and DiCaprio CJ (2010) A multiscale approach to estimating topographically correlated propagation delays in radar interferograms. *Geochemistry, Geophysics, Geosystems* 11.

Lofgren JS, Bjorndahl F, Moore AW, Webb FH, Fielding EJ, and Fishbein EF (2010) Tropospheric correction for InSAR using interpolated ECMWF data and GPS Zenith Total Delay from the Southern California Integrated GPS Network. In: *IGARSS 2010 – 2010 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4503–4506.

Lohman RB and Simons M (2005) Some thoughts on the use of InSAR data to constrain models of surface deformation. *Geochemistry, Geophysics, Geosystems* 6. http://dx.doi.org/10.1029/2004GC00084.

Lohman RB, Simons M, and Savage B (2002) Location and mechanism of the Little Skull Mountain earthquake as constrained by satellite radar interferometry and seismic waveform modeling. *Journal of Geophysical Research* 107: 2118.

López-Quiroz P, Doin M-P, Tupin F, Briole P, and Nicolas J-M (2009) Time series analysis of Mexico City subsidence constrained by radar interferometry. *Journal of Applied Geophysics* 69: 1–15.

Lundgren P, Casu F, Manzo M, et al. (2004) Gravity and magma induced spreading of Mount Etna volcano revealed by satellite radar interferometry. *Geophysical Research Letters* 31: L04602. http://dx.doi.org/10.1029/2003GL018736.

Lundgren P, Usai S, Sansosti E, et al. (2001) Modeling surface deformation observed with synthetic aperture radar interferometry at Campi Flegrei caldera. *Journal of Geophysical Research* 106: 19355–19366. http://dx.doi.org/10.1029/2001JB000194.

Lyons S and Sandwell D (2003) Fault creep along the southern San Andreas from interferometric synthetic aperture radar, permanent scatterers, and stacking. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB001831.

Massonnet D and Feigl KL (1995) Discrimination of geophysical phenomena in satellite radar interferograms. *Geophysical Research Letters* 22: 1537–1540.

Massonnet D, Rossi M, Carmona C, et al. (1993) The displacement field of the landers earthquake mapped by radar interferometry. *Nature* 364: 138–142.

McGuire JJ and Segall P (2003) Imaging of aseismic fault slip transients recorded by dense geodetic networks. *Geophysical Journal International* 155: 778–788. http://dx.doi.org/10.1111/j.1365-246X.2003.02022.x.

Menke W (1989) In: *Geophysical data analysis: Discrete inverse theory*. *International Geophysics Series*, vol. 45. San Diego, CA: Academic Press (revised edition).

Mesinger F, DiMego G, Kalnay E, et al. (2006) North American regional reanalysis. *Bulletin of the American Meteorological Society* 87: 343–360.

Meyer F (2010) A review of ionospheric effects in low-frequency SAR: Signals, correction methods, and performance requirements. In: *IGARSS 2010 – 2010 IEEE International Geoscience and Remote Sensing Symposium*, pp. 29–32.

Michel R, Avouac JP, and Taboury J (1999a) Measuring near field coseismic displacements from SAR images: Application to the Landers earthquake. *Geophysical Research Letters* 26: 3017–3020.

Michel R, Avouac JP, and Taboury J (1999b) Measuring ground displacements from SAR amplitude images: Application to the Landers earthquake. *Geophysical Research Letters* 26: 875–978.

Monti Guarnieri A and Prati C (1996) ScanSAR focusing and interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 34: 1029–1038.

Moore R, Classen J, and Lin YH (1981) Scanning spaceborne synthetic aperture radar with integrated radiometer. *IEEE Transactions on Aerospace and Electronic Systems* 17: 410–420.

Neumann M, Ferro-Famil L, and Reigber A (2008) Multibaseline polarimetric SAR interferometry coherence optimization. *IEEE Geoscience and Remote Sensing Letters* 5: 93–97. http://dx.doi.org/10.1109/LGRS.2007.908885.

Onn F and Zebker HA (2006) Correction for interferometric synthetic aperture radar atmospheric phase artifacts using time series of zenith wet delay observations from a GPS network. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2005JB004012.

Peltzer G, Crampe F, Hensley S, and Rosen P (2001) Transient strain accumulation and fault interaction in the Eastern California Shear Zone. *Geology* 29: 975–978.

Peltzer G, Crampe F, and King G (1999) Evidence of nonlinear elasticity of the crust from the Mw 7.6 Manyi (Tibet) earthquake. *Science* 286: 272–276.

Pepe A, Bertran Ortiz A, Lundgren PR, Rosen PA, and Lanari R (2011) The stripmap ScanSAR SBAS approach to fill gaps in stripmap deformation time series with ScanSAR data. *IEEE Transactions on Geoscience and Remote Sensing* 49: 4788–4804.

Price EJ and Sandwell DT (1998) Small-scale deformations associated with the 1992 Landers, California, earthquake mapped by synthetic aperture radar interferometry phase gradients. *Journal of Geophysical Research* 103: 27001.

Pritchard ME, Ji C, and Simons M (2006) Distribution of slip from 11 $M_w > 6$ earthquakes in the northern Chile subduction zone. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2005JB004013.

Pritchard ME and Simons M (2006) An aseismic slip pulse in northern Chile and along-strike variations in seismogenic behavior. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2006JB004258.

Puyssegur B, Michel R, and Avouac JP (2007) Tropospheric phase delay in InSAR estimated from meteorological model and multispectral imagery. *Journal of Geophysical Research* 112. http://dx.doi.org/10.1029/2006JB004352.

Raney RK (1971) Synthetic aperture imaging radar and moving targets. *IEEE Transactions on Aerospace and Electronic Systems* AE S-7: 499–505.

Remy D, Bonvalot S, Briole P, and Murakami M (2003) Accurate measurements of tropospheric effects in volcanic areas from SAR interferometry data: Application to Sakurajima volcano (Japan). *Earth and Planetary Science Letters* 213: 299–310.

Rodríguez E and Martin JM (1992) Theory and design of interferometric synthetic-aperture radars. *Proceedings of the IEEE* 139: 147–159.

Rosen PA, Hensley S, Joughin IR, et al. (2000) Synthetic aperture radar interferometry. *Proceedings of the IEEE* 88: 333–382.

Rosen PA, Hensley S, Peltzer G, and Simons M (2004) Updated repeat orbit interferometry package released. *Eos, Transactions of the American Geophysical Union* 85: 47. http://dx.doi.org/10.1029/2004EO050004.

Rosen PA, Hensley S, Zebker HA, Webb FH, and Fielding EJ (1996) Surface deformation and coherence measurements of Kilauea Volcano, Hawaii, from SIR-C radar interferometry. *Journal of Geophysical Research* 268: 1333–1336.

Sandwell D, Mellors R, Tong X, Wei M, and Wessel P (2011) Open radar interferometry software for mapping surface Deformation. *Eos, Transactions of the American Geophysical Union* 92: 234.

Schmidt DA and Burgmann R (2003) Time-dependent land uplift and subsidence in the Santa Clara valley, California, from a large interferometric synthetic aperture radar data set. *Journal of Geophysical Research* 108: 2416–2428.

Segall P and Matthews M (1997) Time dependent inversion of geodetic data. *Journal of Geophysical Research* 102: 22391–22409.

Shirzaei M (2013) A wavelet-based multitemporal DInSAR algorithm for monitoring ground surface motion. *IEEE Geoscience and Remote Sensing Letters* 10: 456–460.

Shirzaei M and Burgmann R (2012) Topography correlated atmospheric delay correction in radar interferometry using wavelet transforms. *Geophysical Research Letters* 39: L01305. http://dx.doi.org/10.1029/2011GL049971.

Shirzaei M and Walter TR (2010) Time-dependent volcano source monitoring using interferometric synthetic aperture radar time series: A combined genetic algorithm and Kalman filter approach. *Journal of Geophysical Research* 115. http://dx.doi.org/10.1029/2010JB007476.

Simons M, Fialko Y, and Rivera L (2002) Coseismic deformation from the 1999 m_w 7.1 Hector Mine, California, earthquake as inferred from InSAR and GPS observations. *Bulletin of the Seismological Society of America* 92: 1390–1402.

Snoeij P, Attema E, Davidson M, et al. (2008) Sentinel-1, the GMES radar mission. In: *IEEE Radar Conference, 2008, RADAR'08* 1–5.

Sorenson H (1980) *Parameter Estimation*. New York: Marcel Dekker Inc..

Tarayre K and Massonnet D (1996) Atmospheric propagation heterogeneities revealed by ERS-1 interferometry. *Geophysical Research Letters* 23: 989–992.

Tomiyasu K (1981) Conceptual performance of a satellite borne wide swath radar. *IEEE Transactions on Geoscience and Remote Sensing* 19: 108–116.

Touzi R and Lopes A (1996) Statistics of the stokes parameters and the complex coherence parameters in one-look and multilook speckle fields. *IEEE Transactions on Geoscience and Remote Sensing* 34: 519–531.

Usai S (1997) The use of man-made features for long time-scale InSAR. In: *Proceedings for the IEEE International Geoscience and Remote Sensing Symposium, Singapore*, pp. 1542–1544.

Usai S (2003) A least squares database approach for SAR interferometric data. *IEEE Transactions on Geoscience and Remote Sensing* 41: 753–760.

Watson KM (2002) Satellite interferometric observations of displacements associated with seasonal groundwater in the Los Angeles basin. *Journal of Geophysical Research* 107: 2074.

Wdowinski S, Amelung F, Miralles-Wilhelm F, Dixon T, and Carande R (2004) InSAR-based hydrology of the Everglades, South Florida. *Proceedings of Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. 2004 IEEE International*, vol. 3, pp. 1870–1873.

Wegmuller U (2003) Potential of interferometry point target analysis using small data stacks. In: *Proceedings of Fringe '03 Workshop, Frascati, Italy* 1–3.

Williams S, Bock Y, and Fang P (1998) Integrated satellite interferometry: Tropospheric noise, GPS estimates and implications for interferometric synthetic aperture radar observations. *Journal of Geophysical Research* 103: 27051–27067.

Wright TJ, Parsons BE, and Lu Z (2004) Toward mapping surface deformation in three dimensions using InSAR. *Geophysical Research Letters* 31: L01607.

Xu CJ, Wang H, Ge LL, Yonezawa C, and Cheng P (2006) InSAR tropospheric delay mitigation by GPS observations: A case study in Tokyo area. *Journal of Atmospheric and Solar-Terrestrial Physics* 68: 629–638.

Zebker HA and Goldstein RM (1986) Topographic mapping from interferometric SAR observations. *Journal of Geophysical Research* 91: 4993–4999.

Zebker HA, Rosen PA, Goldstein RM, Gabriel A, and Werner CL (1994) On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake. *Journal of Geophysical Research* 99: 19617–19634.

Zebker HA, Rosen PA, and Hensley S (1997) Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps. *Journal of Geophysical Research* 102: 7547–7563.

Zebker HA, Rosen PA, Hensley S, and Mouganis-Mark P (1996) Analysis of active lava flows on Kilauea Volcano, Hawaii, using SIR-C radar correlation measurements. *Geology* 24: 495–498.

Zebker HA and Villasenor J (1992) Decorrelation in interferometric radar echoes. *IEEE Transactions on Geoscience and Remote Sensing* 30: 950–959.

Zhu X and Bamler R (2010) Very high resolution spaceborne SAR tomography in urban environment. *IEEE Transactions on Geoscience and Remote Sensing* 48: 4296–4308.

This page intentionally left blank

3.13 Geodetic Imaging Using Optical Systems

J-P Avouac and S Leprince, California Institute of Technology, Pasadena, CA, USA

© 2015 Elsevier B.V. All rights reserved.

3.13.1	**Introduction**	387
3.13.2	**Principles**	388
3.13.2.1	Problem to be Solved	388
3.13.2.2	Measurement Principle	389
3.13.3	**Background Information on Optical Sensing Systems**	391
3.13.3.1	Geodetic Scanning Laser	391
3.13.3.2	Passive Optical Imaging	394
3.13.4	**Matching Techniques**	395
3.13.4.1	From 3-D to 2-D Matching	395
3.13.4.2	Algorithms for 3-D Matching	397
3.13.4.3	Algorithms for 2-D Matching	398
3.13.4.3.1	Homogeneous rotation and heterogeneous translation	398
3.13.4.3.2	Optical flow	399
3.13.4.3.3	Statistical correlation	399
3.13.4.3.4	Phase correlation	400
3.13.4.3.5	Regularized solutions and large displacements	400
3.13.5	**Geometric Modeling and Processing of Passive Optical Images**	400
3.13.5.1	The Orthorectification	400
3.13.5.1.1	The orthorectification mapping	401
3.13.5.1.2	Resampling the image	401
3.13.5.2	Bundle Adjustment	402
3.13.5.3	Stereo Imaging	402
3.13.5.4	Processing Flowchart	403
3.13.5.5	Performance, Artifacts, and Limitations	404
3.13.6	**Applications to Coseismic Deformation**	406
3.13.6.1	Usefulness of Coseismic Deformation Measurement from Image Geodesy	406
3.13.6.2	Surface Displacement in 2-D due to the 1999 Mw 7.6 Chichi Earthquake, Measured from SPOT Images	408
3.13.6.3	Surface Displacement in 2-D due to the 2005 Mw 76 Kashmir Earthquake, Measured from ASTER Images	410
3.13.6.4	Surface Displacement in 2-D due to the 1999 Mw 7.1 Hector Mine Earthquake Measured from SPOT Images	411
3.13.6.5	Surface Displacement in 2-D due to the 1999 Mw 7.1 Hector Mine Earthquake Measured from Air Photos	413
3.13.6.6	Surface Displacement in 3-D due to the 2010 Mw 7.2 El Mayor–Cucapah Earthquake from LiDAR and Optical Images Stereomatching	414
3.13.7	**Applications to Geomorphology and Glacier Monitoring**	415
3.13.7.1	Glacier Monitoring	415
3.13.7.2	Earthflows	417
3.13.7.3	Dune Migration	420
3.13.8	**Conclusion**	420
References		422

3.13.1 Introduction

It was realized soon after its invention that photography could be used onboard airborne platforms, initially kites and balloons, for topographic surveying. The first experiment, inspired by the work of mathematician François Arago on image geometry, was actually carried out by Aimé Laussedat in 1849, laying the foundations for photogrammetry (Laussedat, 1854, 1859). There is nowadays a vast archive of photographs taken from various types of aircrafts and spacecrafts available from various national and international agencies and commercial companies. The archive is growing fast as numerous Earth-observing systems or air photo topographic programs are delivering images with ground resolution down to 50 cm or better. Similarly, it also did not take long after the laser was invented by the end of the 1950s before it was used for geodesy and terrain mapping. Over the last decade, systems that operate a laser scanning of the Earth's surface have emerged as new powerful optical systems to sense the Earth's surface from the ground and from airborne and spaceborne platforms (e.g., Carter et al., 2007; French, 2003; Slatton et al., 2007). Geodetic laser scanning is generally referred to as light detection and ranging (LiDAR) or airborne laser swath mapping. Although they differ in fundamental ways, passive and active optical sensing systems both measure a signal reflected at the Earth's surface toward a collector and focused on some sensor. These data provide information on the geometry and physical properties of the Earth's surface. The surface of the Earth is

continuously evolving as a result of geodynamic, climatic, environmental, and human factors. Time series of optical remote sensing data can then in principle be used to monitor those changes and investigate the processes at their origin.

Data collected from optical remote sensing systems can effectively be used to measure the evolution of the topographic surface with enough accuracy to allow investigation of a variety of processes. Probably one of the earliest applications of this approach has been the measurement of ice flow velocity from tracking features such as crevasses or debris on a series of aerial and satellite images (Brecher, 1986; Lucchitta and Ferguson, 1986), prompting later efforts to develop automatic procedures (Scambos et al., 1992). Application of this approach to the solid Earth, for example, the measurement of ground displacements or topographic changes induced by earthquakes and geomorphic processes, has been explored in a number of studies (e.g., Aryal et al., 2012; Corsini et al., 2009; Crippen, 1992; Delacourt et al., 2007; Kääb et al., 1997; Mackey et al., 2009; Oskin et al., 2012; Roering et al., 2009; Van Puymbroeck et al., 2000). This is a rapidly growing area of research due to the need for a better understanding of those processes, motivated in particular by the need to monitor and understand better the impact of climate change on the landscape and water resources and the growing body of optical remote sensing techniques (Bishop et al., 2012; Tarolli et al., 2009).

In this chapter, we review the methods used in such studies and illustrate with particular applications the measurement of the Earth's surface changes produced by earthquakes, ice flow, landsides, and sand dune migration. We are concerned here with characterizing the geometric changes of the Earth's surface, which have occurred between various epochs of acquisition. We do not cover the literature on the characterization of tectonic or geomorphic processes from morphometric measurements. The reader is referred to textbooks or review papers on that topic (Burbank and Anderson, 2001; Kirby and Whipple, 2012). We focus on optical remote sensing systems but many of the techniques described here apply to radar images. With regard to the processing and exploitation of radar images, the reader is referred to review papers on this technique (Massonnet and Feigl, 1998) and to the chapter by Simons and Rosen in this same volume (**Chapter 3.12**).

Here, we start with describing the principle of how optically sensed terrain models might be used to quantify geologic and geomorphological processes. We next move on with describing practical implementations for the exploitation of either LiDAR or optical images. We do not review in-depth photogrammetric and LiDAR techniques, but we mention the aspects of importance with regard to applications to Earth sciences. We illustrate the potential and limitations of these techniques based on an overview of case studies, and finally, we discuss research perspectives.

3.13.2 Principles

3.13.2.1 Problem to be Solved

We are interested in quantitatively characterizing geometric changes of the Earth's surface topography between two epochs of acquisition of remote sensing optical data (**Figure 1**). Let us consider that the data provide some rendering of the Earth's surface, S_1 and S_2 at time t_1 and t_2, respectively. S_i would characterize the topographic surface at time t_i with respect to a geodetic reference frame (represented by Oxy in **Figure 1** where the problem is sketched in 2-D), which in practice could be a particular realization of the International Terrestrial Reference Frame (Altamimi et al., 2002) associated to a particular datum and its optical reflective properties. The optical properties of the surface and the geometry contribute to determining the radiometry measured by any optical system, whether passive or active. The surface cover (vegetation and human infrastructures) and the substrate determine these properties.

In practice, the geometry is represented by a digital elevation model (DEM), which is a discretized representation of the topography elevation. The sampling grid can be regular or not depending on the technique used.

Let us now consider a material point in the subsurface located at M_1 at epoch t_1. This same point lies at M_2 at epoch t_2, with respect to the same geodetic reference frame. In practice, the displacement vector (d_x, d_y, d_z) could be due to tectonics (e.g., an earthquake) or other processes such as landslide or ice flow as we will see in the application section

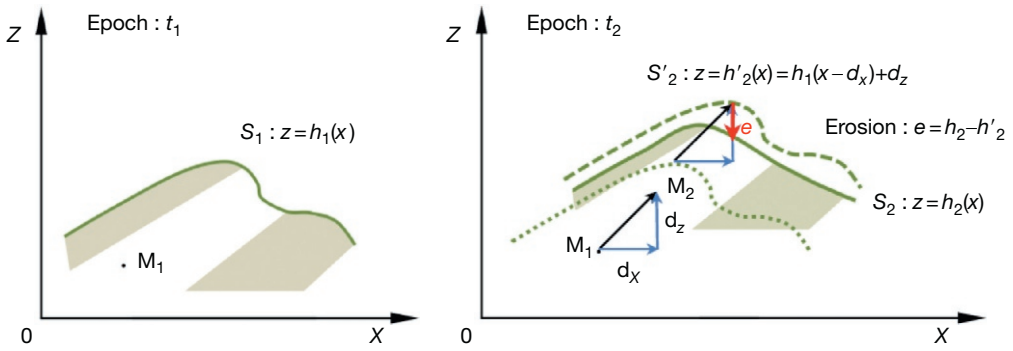


Figure 1 Schematic 2-D representation of the problem to solve. We are interested in characterizing quantitatively geometric changes of the Earth's surface topography between epochs t_1 and t_2. S_1 and S_2 refer to the Earth's surface at time t_1 and t_2, respectively, with respect to a geodetic reference frame, represented by Oxy. The Earth's surface may have changed as a result of displacement of the subsurface medium with respect to the reference frame, represented by the displacement vector M_1M_2. It may also have changed as a result of erosion or sedimentation at the Earth's surface, represented by the scalar quantity e.

in the succeeding text. Note that the medium around M_1 could have deformed although this is not represented in Figure 1 for the sake of simplicity. The surface topography, represented by S_1 and S_2 at epochs 1 and 2, respectively, is not a passive marker in general. Between epochs t_1 and t_2, it may have evolved as a result of erosion or sedimentation. As a result, advective transport of the initial topography yields a surface S'_2, which differs from the topography at epoch t_2 (dashed line in Figure 1). The elevation difference between S'_2 and S_2 is the measurement that quantifies the evolution of the topography due to erosion ($h_2 < h'_2$) or sedimentation ($h_2 > h'_2$). The applications reviewed in this chapter hinge on the measurements of either topographic changes, that is, the difference between h_2 and h'_2, a scalar field $e = h_2 - h'_2$, or the ground displacement vector field (d_x, d_y, d_z) (represented by M_1M_2 in Figure 1). The information derived from sensing the Earth's surface with optical systems is, however, inherently insufficient to solve for both the change of elevation of the topography and the ground displacement vector. In principle, any erosion or sedimentation should make it impossible to decompose the measured difference in elevation between two epochs into a ground displacement (d_x, d_y, d_z) and erosion of the topography. Thus, in practice, one or the other term must be assumed negligible or known independently. Often, these assumptions appear natural given the context of the observations and are not always stated explicitly.

In most geodetic applications, it is assumed that the topography is advected as a passive marker. It follows that the displacement field, which allows matching the topography at epochs t_1 and t_2, would also be matching the radiometric texture of the surface, provided that it has not changed between the two epochs (as they might have due to change of the land cover, surface hydrology, or human activities). Let us stress here the fundamental difference between the 'geodetic optical imaging' techniques described in this chapter and standard geodetic techniques, which allow measuring directly the displacements of material points M_1M_2.

3.13.2.2 Measurement Principle

Let us assume that the same portion of the Earth's surface was sensed at two epochs from optical sensing methods and that these data were used to produce perfectly registered DEMs and some representation of surface optical properties at the two epochs. Let us refer to h_1 and h_2 as the functions describing the topographic surface at time t_1 and t_2, respectively. The DEMs are discrete sampling of these functions. In general, the elevation change measured from the difference between the topographic surfaces, $h_2 - h_1$, will combine the effect of advection and erosion, $e(x,y)$, of the surface (Figure 1):

$$h_2(x,y) - h_1(x,y) = h_1(x - d_x, y - d_y) + d_z(x,y) - h_1(x,y) + e(x,y) \quad [1a]$$

The term in the brackets on the right side represents the elevation change due to horizontal advection of the topography. Assuming that this equation can be approximated by a Taylor expansion to first order, we get

$$h_2(x,y) - h_1(x,y) \approx d_z(x,y) - d_x(x,y)\frac{\partial h_1}{\partial x}(x,y) - d_y(x,y)\frac{\partial h_1}{\partial y}(x,y) + e(x,y) \quad [1b]$$

If ground displacements can be neglected, changes of the topography are most simply characterized by differencing the two topographic surfaces:

$$e = h_2 - h_1 \quad [2]$$

This yields directly an estimate of erosion ($e < 0$) or sedimentation ($e > 0$) at the Earth's surface (Figure 2). As erosion and sedimentation presumably reset the surface optical properties, this measurement is in principle the only one that is meaningful in the presence of erosion or sedimentation. The measurement requires essentially some technique to resample the two DEMs on a common grid. This resampling procedure should in principle take into account how the DEMs were produced so as to respect the physics of the measuring technique. Resampling errors will inevitably be introduced.

In practice, DEMs produced independently from the data acquired at different epochs are not perfectly registered. As a result, differencing the topographic surfaces may, for a large part, reflect the resulting bias with registration errors possibly in excess of the signal of interest. Figure 1 can be taken to illustrate this issue if (d_x, d_y, d_z) is now meant to represent a misregistration (ϵ_x, ϵ_y, ϵ_z). Due to the bias introduced by the misregistration, eqn [2] becomes

$$h_2(x,y) - h_1(x,y) = e(x,y) + h_1(x - \epsilon_x, y - \epsilon_y) - h_1(x,y) \quad [3a]$$

or in its Taylor expansion form

$$h_2(x,y) - h_1(x,y) \approx e(x,y) + \epsilon_z(x,y) - \epsilon_x(x,y)\frac{\partial h_1}{\partial x}(x,y) - \epsilon_y(x,y)\frac{\partial h_1}{\partial y}(x,y) \quad [3b]$$

The data analysis then requires a procedure for precise co-registration of the DEMs so as to minimize this bias. In most instances, the co-registration will be achieved by assuming that some particular areas have not experienced any topographic changes (the topographic differences in those areas should be null) or by using a priori constraints on the displacements at some ground control points (GCP). In general, DEM differencing will therefore reflect the combined effects of misregistration, resampling errors, and advective transport of the topography.

If the topography is assumed to have been transported advectively, in which case $e = 0$ (Figure 3), the 3-D displacement field between two epochs might in principle be retrieved from matching the two DEMs. This requires some technique for matching DEMs in 3-D. The matching procedure solves for the displacement field vector (d_x, d_y, d_z), which satisfies

$$h_2(x,y) = h_1(x - d_x, y - d_y) + d_z(x,y) \quad [4a]$$

or in its Taylor expansion form

$$h_2(x,y) - h_1(x,y) \approx d_z(x,y) - d_x(x,y)\frac{\partial h_1}{\partial x}(x,y) - d_y(x,y)\frac{\partial h_1}{\partial y}(x,y) \quad [4b]$$

Equation [4b] illustrates that the determination of the displacement field from matching the topography measured at

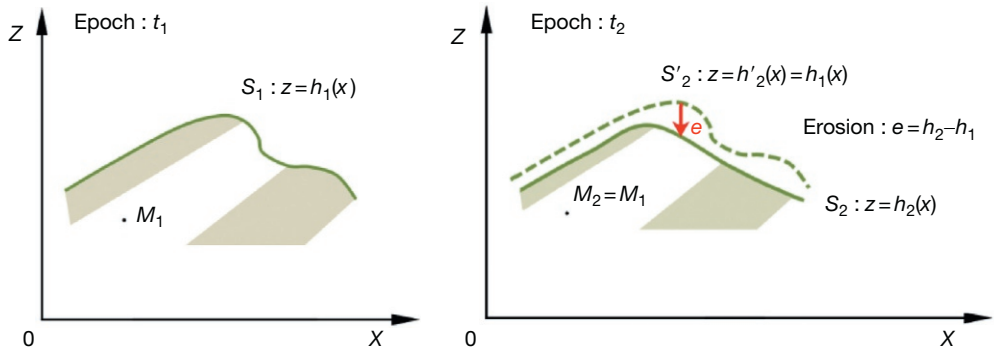


Figure 2 Simplified version of **Figure 1** in the case with no advective transport of the subsurface ($M_1 = M_2$). The Earth's surface may have changed as a result of erosion or sedimentation represented by e.

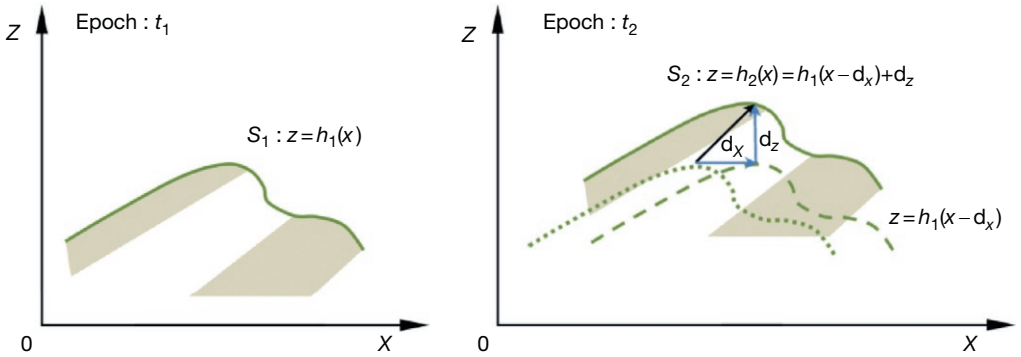


Figure 3 Simplified version of **Figure 1** in the case with no erosion nor sedimentation. The Earth's surface is simply advected according to ground displacement vector field M_1M_2.

two epochs is intrinsically an ill-posed problem: only the displacement along the gradient of the topography can be determined. Some assumptions are therefore needed regarding the regularity of the displacement field.

A simple procedure to regularize the matching problem (whether the quantity to be matched is the topography or any other scalar field) is to assume that the displacement field is continuous and varies smoothly (continuously differentiable). In that case, the horizontal displacement vector at a given point M_1 can be determined from optimizing the matching between two windows of the same size w centered on M_1 in h_1 and on M_2 in h_2 as a function of the position M_2 (**Figure 3**). For the regularization to be effective, the window size must be large enough so that the direction of topographic gradient varies significantly within that window. In practice, this requires the window size to be at least five to ten times larger than the average distance between measurements. The measurement provides an estimate of some average of displacement within that window. The nature of the averaging depends on the choice of a particular matching procedure. An important implication is that the displacement field is always resolved with a lower spatial resolution than the original DEM. In principle, regularization can be achieved with a relatively small matching window, 3×3, for example, for a scene rich in small-scale features of various orientations. Images from natural scenes generally require larger windows. The spatial resolution is therefore generally no better than about five times the ground sampling distance (GSD) of the two stereoscopic optical images.

In the case of DEMs obtained from geodetic laser scanning, some radiometric information (the reflected intensity or the waveform of the reflected pulse) might be available in addition to the geographic coordinates of the scanned points. This information can in principle be used to optimize and help regularize the matching problem (again assuming the advective transport of the topography). This requires that the radiometric measurement can be converted into a stationary property of the ground's surface. In practice, the effects of the atmosphere and changes of the land cover can be a limitation.

In the case where the DEMs were computed from stereoscopic pairs of images, matching of DEMs is, however, not an optimal approach. This is so because DEMs generally fall short of representing accurately the information contained in the original data used to construct the topography. The radiometry at one pixel of an optical image depends on the surface's optical properties (determined by the land cover and substrate) and local topography (terrain roughness and average slope at the scale of the sensed spot on the ground), modulated by the atmospheric filter and transfer function of the optical system. If this texture is advectively transported with the topography, it is then a richer source of information on ground displacement than the DEM itself. It follows that ground displacements can

be measured more accurately by matching the image texture, very much the same way parallax offsets are measured to calculate DEMs. In addition, from a mathematical point of view, matching the radiometry and matching the topography are equivalent problems, both ill-posed. However, the topographic and radiometric gradients do not need to be parallel so that when radiometry is used, the determination of the displacement field is in principle less of an ill-posed problem. Note, however, that this is not true if the surface has a uniform albedo as the radiometry will then be entirely determined by the topography (e.g., sand dunes). Regularization of the matching problem is less stringent as the optical images generally have more texture than the DEMs at high spatial frequencies, simply because the DEMs are themselves often produced from the determination of stereoscopic offsets from matching the images. The regularization of the matching problem imposes that the DEMs have a lower spatial resolution than the images they were derived from.

The measurement of surface displacements from optical remote sensing data or directly from DEMs thus relies on matching measurements acquired at different epochs. In essence, matching techniques yield at any point of a reference dataset, a measurement of the vector field that best brings into coincidence a window centered on that point with a corresponding window in the second dataset. The output of the matching procedure is a vector field (**Figures 4** and **5**), which can be represented by shaded representation of the horizontal and vertical components in 3-D, as will be the case in the studies shown in the succeeding text.

As an illustration, **Figure 6** shows the output from matching two Satellite Pour l'Observation de la Terre (SPOT) images, with a ground resolution of 10 m, acquired before and after the 1999 Mw 7.1 Hector Mine earthquake (Leprince et al., 2007). These images were orthorectified, to remove stereoscopic distortions due to the topography, and correlated using the methodology described in the succeeding text and a preexisting regional DEM. In this case, the offset field between the two orthoimages should show both horizontal displacements due to the earthquake and orthorectification errors. Clearly, the offset field is dominated by the ground displacement induced by the earthquake: the surface rupture shows up as a discontinuity of surface displacements. Profiles that run across the fault trace can be used to measure surface fault slip with accuracy better than 1 m (1/10 of pixel size). In this particular case, the offsets due to inaccurate modeling of stereoscopic effects related to the topography are small compared to the amplitude of the displacement signal. In most studies, reduction of these artifacts is a critical challenge.

In principle, the approach outlined here might be applied to passive or active optical data. In both cases, the exploitation of optical remote sensing data for documenting geometric changes of the Earth's surface requires some model of the imaging system, namely, a model that allows projecting back on the surface the information collected by the optical sensor. Recent advances in geodetic imaging from optical methods have, for a great deal, resulted from the improved accuracy of this geometric modeling. In the following section, we provide background information on optical sensing systems relevant to the development of such models.

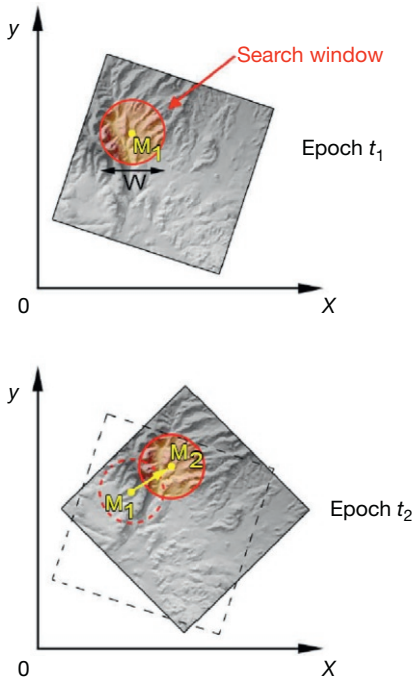


Figure 4 Scheme of the matching procedure used to determine offsets between two datasets (two images or two DEMs and two clouds of LiDAR data). Because the matching (eqn [4]) is intrinsically ill-posed, the offset vector is determined from optimizing the matching between a window centered on a running point M_1 in the dataset acquired at epoch t_1 and a search window of same size in the second dataset. The window size must be large enough that it contains enough texture to solve for the matching problem. The measured offset is the vector M_1M_2, where M_2 is the center of the best matching position of the search window in the second dataset. The principle holds in 2-D, as represented here, and in 3-D.

3.13.3 Background Information on Optical Sensing Systems

A key element in image geodesy, as with photogrammetry, is the proper modeling of the imaging system so that the accuracy of the projection on the Earth's surface of the signal measured by the optical sensing system meets geodetic standards. The modeling is specific to the particular imaging system. It is therefore important that the users be aware of the various elements that determine this geometric modeling and the various potential factors of geometric distortions. These distortions introduce misregistrations that, if not compensated for, will bias the measurement of topographic changes. The focus of this section is therefore to introduce those factors in the case of both active optical sensing and passive optical sensing.

3.13.3.1 Geodetic Scanning Laser

It was not long after the laser was invented at the end of the 1950s that it started being used from space to sense the surface of Earth satellites and other planets (Arnold, 1967; Kovalevsky and Barlier, 1967). The principle of the geodetic laser scanning is simple (e.g., Baltsavias, 1999; Carter et al., 2007) (**Figure 7**). The laser technology allows the production of short intense pulses of monochromatic light. A variety of instruments that

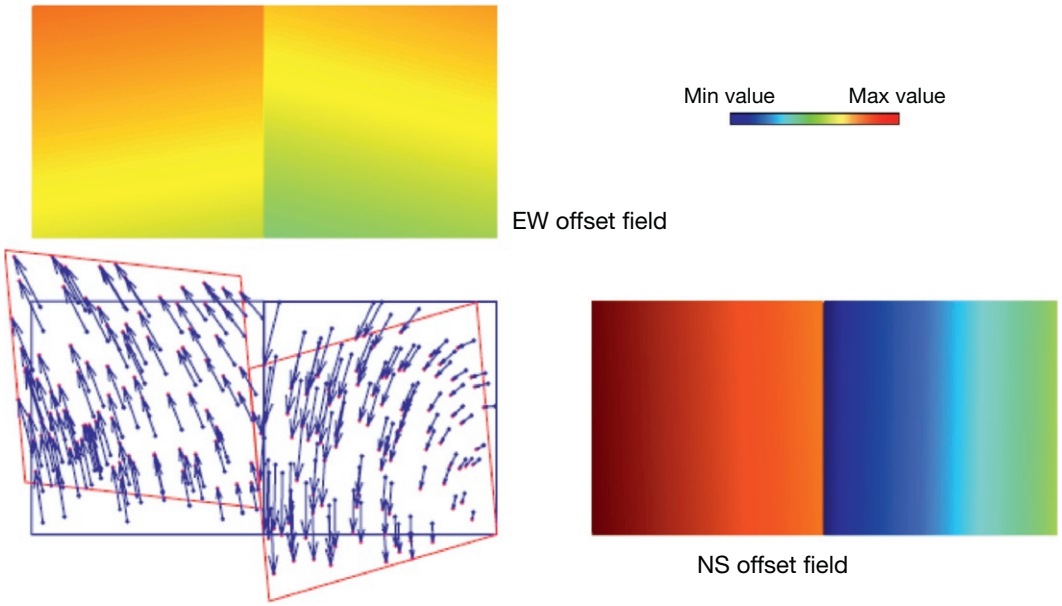


Figure 5 The output of the matching procedure between two datasets is an offset field, which can be represented as a vector field or as scalar fields for individual components. Offset can be measured in the space or image space. In the space domain, those offset would represent ground displacement and potential misregistrations of the dataset. The figure represents schematically the horizontal displacement field due to an earthquake corrupted by registration errors.

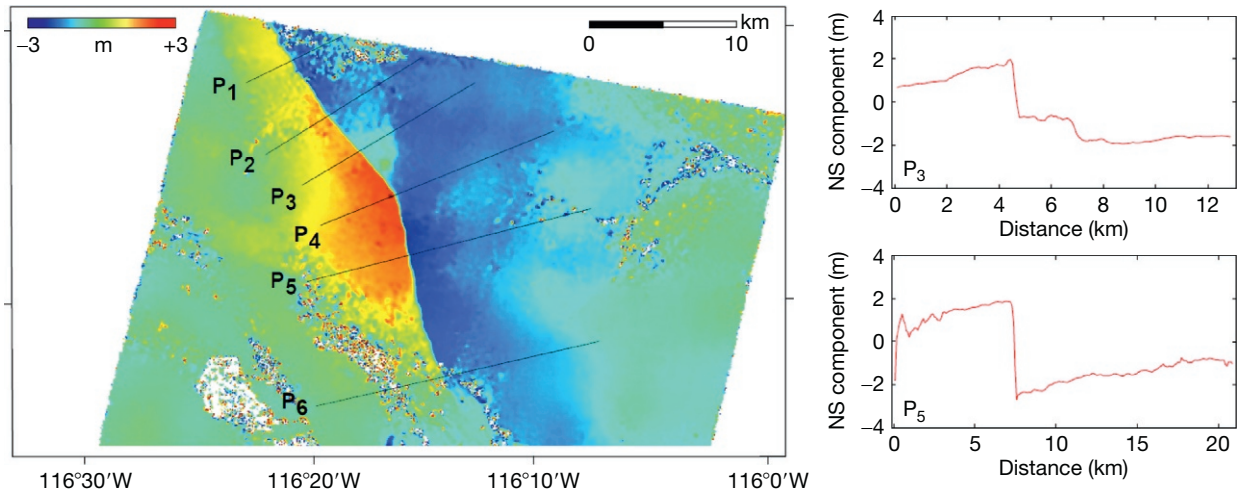


Figure 6 North component (positive to the North) of the coseismic displacement field due to the 1999 Mw 7.1 Hector Mine earthquake in California measured from correlated SPOT 2 and SPOT 4 monochromatic images with 10 m GSD acquired on 12 August 1998 and 10 August 2000. Both images were orthorectified and co-registered on a 10 m-resolution grid using COSI-Corr. Offsets were measured from subpixel correlation with a 32×32-pixel sliding window and a 16-pixel step. The offset field was denoised using the nonlocal means filter (Buades et al., 2008). The standard deviation on individual measurements is around 0.8 m. Right panels show 2 km wide swath profiles across the fault trace. These profile showing a clear discontinuity of surface displacement at the fault trace with up to 5.5 m right-lateral strike slip.

differ with regard to the energy per pulse, the number of pulses per second, and the electro-optical scanning system can be operated from the ground, aircraft, or space platforms. Lasers for airborne geodetic applications generate 5–10 ns long pulses with a frequency 50–150 kHz at a wavelength in the infrared (1064 nm, e.g., for neodymium-doped yttrium aluminum garnet) (e.g., Carter et al., 2007). The narrow bandwidth allows tight collimation of a beam, which is deflected toward the target using an oscillating mirror. An optical system collects, filters, and focuses the reflected pulse on a photodetector. The two-way travel time is measured and provides a determination of the range. The uncertainty in the measured laser range results from the uncertainties on flight time, atmospheric correction, and range walk. The 1-σ precision is typically 2–3 cm in airborne surveys. The intensity of the reflected light depends on the distance to the reflecting surface (it decays at

$1/d^2$), orientation, and optical properties at the given wavelength (reflectance and roughness). The measurement of the direction of the beam relative to the sensor together with the flight time provides information on the position of the reflecting surface, averaged over the spot size, relative to the sensor. This information about the position of the reflecting surfaces with respect to the sensor (the 'interior orientation model') is combined with the information about orientation of the sensor determined from the navigation (the 'exterior orientation model').

The orientation of the sensor head is measured from an inertial measurement unit (IMU), and its position is measured from GPS receivers on board the aircraft. The position of these dual-frequency receivers is determined, at a sampling rate of about 5 Hz, from kinematic GPS processing relative to a set of GPS stations on the ground. The exterior orientation model therefore consists of six measurements (the roll, pitch, and yaw characterizing the pointing direction of the sensor head and its geographic coordinates in 3-D) (**Figure 7**).

The infrared beam can be reflected by the vegetation as well as the ground surface. Multiple reflections from the canopy and ground surface are generally detected over vegetated areas. Postprocessing is then needed to separate reflections from the ground and canopy.

The interest of airborne LiDAR for geographic mapping was explored early on (Krabill et al., 1984) and later on, as the technique became more affordable, for geomorphic and seismotectonic applications (e.g., Hudnut et al., 2002; McKean and Roering, 2004; Woolard and Colby, 2002). Much effort has been made over the last decade to collect airborne LiDAR data over areas of potential interest for tectonics and geomorphology in particular thanks to the establishment in 2003 of the National Center for Airborne Laser Mapping (http://www.ncalm.cive.uh.edu/). Existing surveys generally consist of shots with a density of a few points per square meter.

Altogether, the technique provides the positions of a cloud of points in 3-D and their associated intensities. The positions are measured relative to the reference frame defined by the positions assigned to the ground-based GPS stations in the kinematic GPS processing, for example, some realization of the ITRF system. Registration errors might result from both the inaccuracies of the interior and exterior geometric models and errors on the positioning of the ground-based stations. Vertical and horizontal errors (at the 1-σ confidence level) are on the order of 5–10 and 10–25 cm, respectively. The main source of error is probably due to the uncertainty on the elevation of the sensor, which can be as large as 15 cm, due to the difficulty of modeling accurately the effect of the troposphere on kinematic GPS (Shan et al., 2007). Intensities

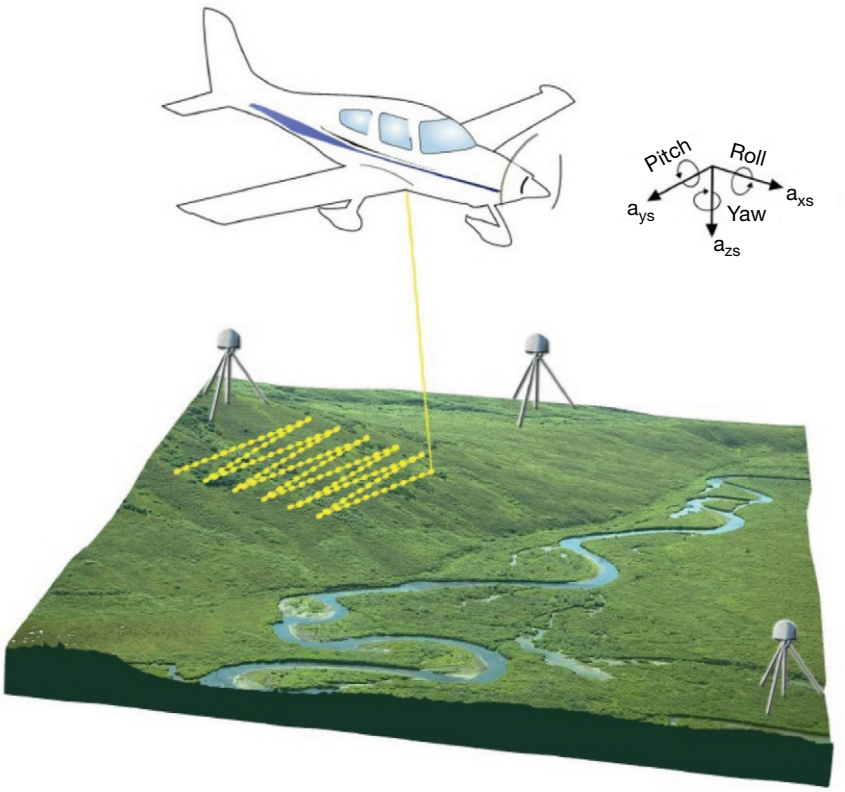


Figure 7 Setting for an airborne LiDAR survey. An optical scanner distributes laser pulses in a zigzag pattern within a swath on the ground. The beam direction relative to sensor, the intensity of the return, and the two-way travel time required for each pulse to travel to and from a reflecting point below are recorded. The position of the sensor is determined from the positions of GPS receivers onboard the aircraft relative to a set of local GPS ground stations. The orientation (roll, pitch, and yaw) and accelerations of the sensor head (*axs*, *ays*, *azs*), measured from an inertial measurement unit, are used, along with the scanner mirror angle and measured range values, to calculate the coordinates of surface points. Modified from Carter WE, et al. (2007) Geodetic laser scanning. *Physics Today* 60 (12): 41–47.

can potentially be exploited to match datasets acquired at different times provided that the geometric attenuation of the return pulse is corrected for and that reflective properties of the ground are stationary.

3.13.3.2 Passive Optical Imaging

In passive imaging systems, the terrain is illuminated by the natural light emitted by the Sun or possibly reflected by the Moon. In this case, only the energy integrated over the bandwidth of the imaging system is measured since the source signal is incoherent.

A passive optical remote sensing system consists of a platform and an optical system to collect the light (telescope), eventually filter it in several spectral bands, and focus each band on detectors (Figure 8). Panchromatic systems measure at each pixel the intensity of the light collected across the visible range. In multispectral imaging, the visible to near-infrared range is filtered into a number of narrowbands. Spectral resolution generally comes at the expense of spatial resolution due to the limited sensitivity of the detectors, the limited storage capacity onboard the platform, and the bandwidth for data downloading. For geodetic applications, there is generally no advantage to using multispectral data. This is because most of the factors on geometric distortions, which limit the application, are common to all the bands and the various bands are generally correlated, so the matching accuracy does not really scale as the inverse of the square root of the number of bands as one might hope.

The platform can be an aircraft or a spacecraft. Its position and orientation are generally estimated from the navigation information of the aircraft or from the information on the orbit and attitude of the spacecraft. This information, together with the orientation of the optical axis of the collector with respect to the platform, defines the exterior orientation model.

The detectors are charge-coupled devices (CCDs), which are organized in linear or 2-D arrays. The 'interior orientation model' defines the position of each CCD in the focal plane of the collector. Generally, the CCD spacing is adapted to the resolving power of the telescope characterized by its point spread function (PSF). Generally, the CCD spacing is about half the width of the PSF so that optical images are generally aliased (to avoid aliasing, the CCD spacing should be about 1/5 of the width of the PSF). The distance between the pixel centers projected on the ground is referred to as the GSD. In principle, this distance varies within an image depending on the topography and geometric model of the optical system. It is typically 15–30 m for Landsat images (for the bands in the visible range), 10 m for SPOT 1 to SPOT 4 panchromatic images, 2.5–5 m for SPOT 5 panchromatic images, 15 m for ASTER images, and 50 cm to 1 m for IKONOS and DigitalGlobe images. The resolution of film-based or digital aerial photographs is generally metric to submetric for standard topographic survey.

The intensity measured at a pixel of a digital image results from the optical properties, roughness, and slope orientation of the spot at the Earth's surface that is contributing to the reflected light collected by the CCD (the size of the spot is determined by the PSF) and from the filtering effect of the atmosphere.

The position of a CCD in the focal plane of the image determines the direction toward the spot on the ground that is sensed by this particular CCD. The unit vector pointing

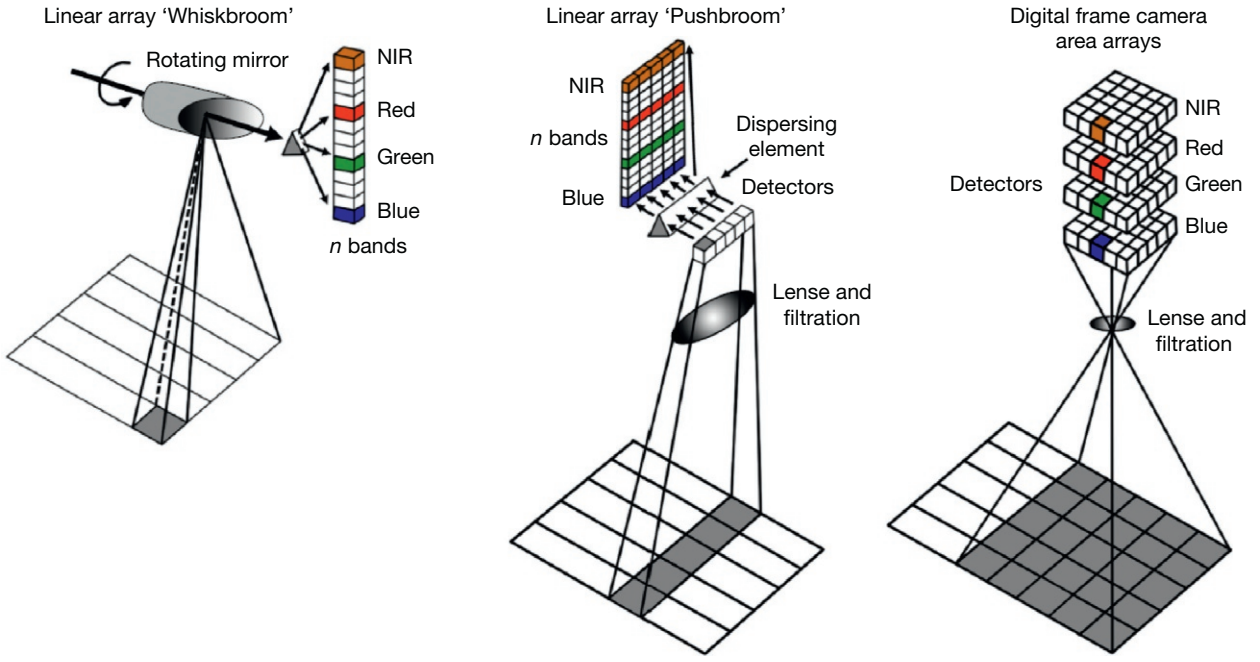


Figure 8 Schematic representation of wiskbroom, pushbroom, and frame camera imaging systems. The gray area shows for each imaging system the pixels measured simultaneously during the image acquisition. The exterior orientation, which defines the orientation of the optical axis of the collector, is common to all these pixels. Modified from Jensen JR (2006) *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River, NJ: Prentice Hall.

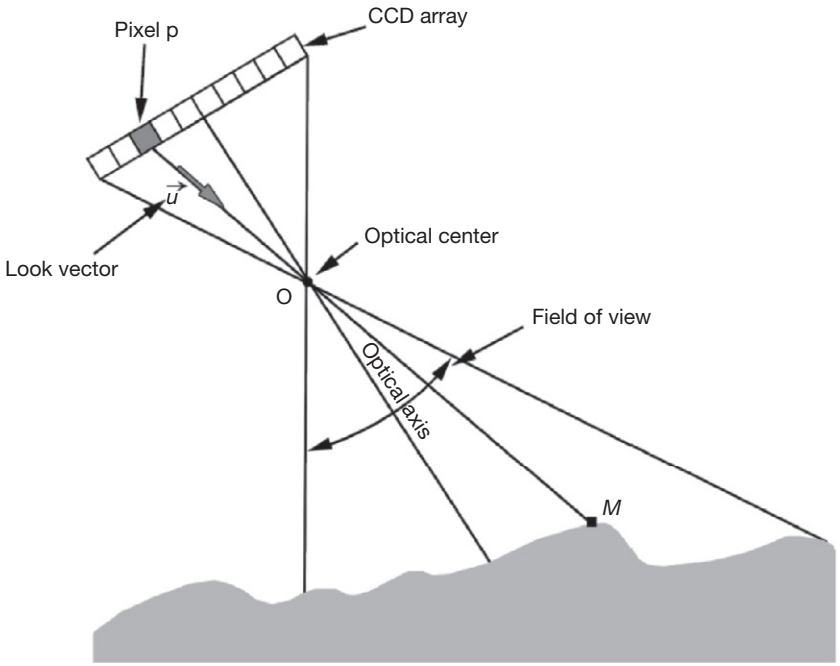


Figure 9 The spot on the ground around point M that illuminates pixel p in the focal plane of the imaging optical telescope is determined based on classical optical geometry. Light is assumed to follow a ray connecting M and p through the optical center of the collector. The position of M relative to p depends on the interior orientation model (where the CCD corresponding to pixel p lies in the focal plane), on the position of the optical center O, and the exterior orientation model (the orientation of optical axis of the collector).

along that direction is called the look vector (Figure 9). The interior orientation model defines its orientation relative to the optical axis of the telescope.

A traditional analogue camera or a digital camera scans the light collected simultaneously within the field of view of the telescope (Figure 9). Only six parameters are necessary to characterize the exterior orientation of such a frame camera at the time of acquisition of a particular image (the geographic position in 3-D of the optical center and the roll, pitch, and yaw of the platform). The interior orientation model is in principle fixed and generally has been calibrated by the manufacturer of the sensor. The calibration model accounts for the geometric distortions due to the aberrations of the telescope, the focal length of the telescope, and the physical position of the CCD in the focal plane. In principle, only six parameters need to be determined to characterize the ground projection of any pixel on the ground. These six parameters determine uniquely, given the interior model, the position of the optical center of the image and the look vector at any point in the image (the three angles determining the orientation of the ray hitting a particular CCD or point of an analogue film). Optimization of the geometric modeling requires reducing the errors on the a priori estimate of only these six parameters. As is customary in photogrammetry, a small number of GCPs may be required as is detailed in the succeeding text.

Optical satellite remote systems take advantage of the satellite motion along its track to scan the ground. Some systems (such as Landsat launched in 1972) operate a whisk broom scanning, similar to the LiDAR scanning system described in the previous section in which only one pixel is sensed at a time. The interior model is determined by the rotating mirror, which allows line scanning. Each pixel is acquired at a different time along the track. It results that each pixel has an independent look vector determined by the attitude of the satellite and orientation of the scanning mirror at the time of light detection. The errors on these look vectors are for a large part independent and cannot be optimized globally.

Most systems, however, operate a push broom scanning in which an entire line of the image is acquired at a given time (Figure 8). In that case, the parameters of the exterior orientation model are common to each line. This is a better situation than for the whisk broom system as the estimated exterior model can be optimized to improve the registration of the image as misregistrations errors due to the exterior orientation model are common for a line. The internal orientation (IO) model is fixed and can be improved using an in-flight calibration procedure (Leprince et al., 2008b).

3.13.4 Matching Techniques

3.13.4.1 From 3-D to 2-D Matching

In essence, matching techniques are meant to yield at any point of a reference space (epoch t_1) a measurement of the offset vector field that best brings into coincidence this point with a paired point in a deformed space (epoch t_2). Matching can in principle be carried on in the 3-D physical space or the 2-D image space. In the physical space, the output will be directly a measurement of the displacement vector, provided that the Earth's surface has been advectively transported between

epochs t_1 and t_2. The noise will come from the misregistration of the data. In the image space, the measured 2-D offset will reflect both stereoscopic effects and ground displacement.

The matching criterion can be based on some radiometric measurement or on the geometry of the sensed surface provided that both can be assumed to have been advected with no or negligible modifications. In case the geometries of the sensed surfaces are matched, there is an implicit assumption that there exists a scale at which the geometry has been preserved. This scale is defined by the size of the search window used in the matching procedure.

As we mentioned earlier, matching the scalar function is intrinsically an ill-posed problem. In principle, matching various bands of a multispectral image should help alleviate the ill-posedness. In practice, this is not that effective due to the strong correlations among the various bands, to the variability of the measured radiometry due to the atmosphere variability, and also to geometric and environmental modifications of the Earth's surface. For this reason, it is also necessary to regularize matching based on the radiometry, generally through matching the radiometric texture within a search window.

Let us now assume that we have a set of optical data, which were acquired at two epochs t_1 and t_2, and accurate geometric models of the imaging systems. Some matching procedure is wished to measure ground displacement and to improve the co-registration of the datasets.

Let us first consider the case where the data consist of a digital rendering of the topography from LiDAR measurements or some other technique, referred to as DEM. The dataset consists of a cloud of points at the Earth's surface with their positions defined in 3-D with respect to some reference frame. The differencing of two DEMs acquired at different epochs is the simplest representation of topographic changes between the two epochs. This operation requires resampling of the dataset on a common grid. Matching the two DEMs might, however, be a more relevant measurement. This is the case if ground displacement has occurred and if surface changes due to erosion, sedimentation, or land cover modifications can be neglected. As mentioned in the preceding text, the matching procedure solves for the displacement field vector (d_x, d_y, d_z), provided regularization assumptions, so that eqn [4] is verified as closely as possible.

Algorithms have been developed in computer vision, which allow 3-D matching of digital representations of surfaces, which include various possible regularization techniques. An example of such an algorithm is the Iterative Closest Point (ICP) technique (Besl and McKay, 1992), which has been tested recently on LiDAR data (Nissen et al., 2012; Teza et al., 2007). Details about this approach and its performance are given in the next section.

As the topography can always be parameterized in 2-D, most simply by expressing elevation as a function of geographic coordinates, the 3-D matching of optical remote sensing data of the Earth's surface can generally be transformed into a 2-D matching problem, except at locations of cliffs and overhangs. In practice, the matching problem expressed in eqn [4] can be solved in two steps as illustrated in **Figure 3**. First, the horizontal displacement fields, $d_x(x,y)$ and $d_y(x,y)$, can be determined from a 2-D matching technique as correlation of h_1 and h_2 is unaffected by the shift represented by the vertical displacement $d_z(x,y)$ (e.g., Aryal et al., 2012; Borsa and Minster, 2012). The vertical displacement field can be obtained next by differencing the topography at epoch 2 and the topography measured at epoch 1 advected horizontally (**Figure 3**). Such a measurement is here also biased by registration errors.

In case of passive optical images, the geometric modeling of the imaging system provides in principle a determination of the look direction at each pixel in the image. If the topography is known independently at epochs t_1 and t_2, the model can be used to produce orthoimages.

The geometric modeling can in principle be optimized and ground displacements retrieved from matching these orthoimages in 3-D. Errors in the DEMs and registration errors of the DEM relative to the images introduce spurious geometric distortions of the orthoimages. These distortions can be systematic and quite large in the common situation where the optical images have a better ground resolution than the DEM or if a DEM is available at only one epoch while the topography is known to have changed (e.g., due to advective transport). The geometric distortions are enhanced at higher ground resolution due to the topographic roughness being proportionally larger. This difficulty seriously limits the benefit of using higher-ground-resolution images to improve the resolution of ground displacement measurements. As a result, 2-D matching of orthoimages is generally not an optimal approach.

In the ideal case where stereoscopic pairs of images are available at epochs t_1 and t_2, it is in principle possible to solve accurately the 3-D matching problem. In that case, the 3-D matching problem can be reformulated as a 2-D matching problem as illustrated in **Figure 10**. The 3-D vector can indeed be retrieved from measuring offsets between images projected on a reference surface, in practice a reference ellipsoid.

In this chapter, we refer to the offset field as the horizontal vector field retrieved from matching two 'images.' The 'image' can be the intensity measured from an optical camera or the elevation.

In the case of optical images of the same ground area taken from different view angles, the offsets measured from matching the images projected on the ellipsoid will represent stereoscopic parallax effect if the images are synchronous or a combination of stereoscopic effect and ground displacement. The most general procedure to measure displacements in 3-D with optical images is therefore measuring offsets using two pairs of stereoscopic images (**Figure 10**). Three independent offset fields can then be derived.

The determination of the respective contribution of stereoscopic effects and ground displacement to these offsets is then simply determined by the geometry (**Figure 10**). For the sake of simplicity in **Figure 10**, the focal points of the imaging system corresponding to all four images are supposed to be coplanar. The intersection of this plane with the reference ellipsoid defines the epipolar direction, the direction of offsets induced by stereoscopic effects. The offsets measured along the perpendicular direction would in principle be free of stereoscopic effects due to the topography and should result only from misregistration and ground displacement. In reality, the four focal points would not be coplanar so that a different epipolar direction is defined for each pair of images.

In principle, the three offset fields, which result in six independent measurements at each point on the ground, can be

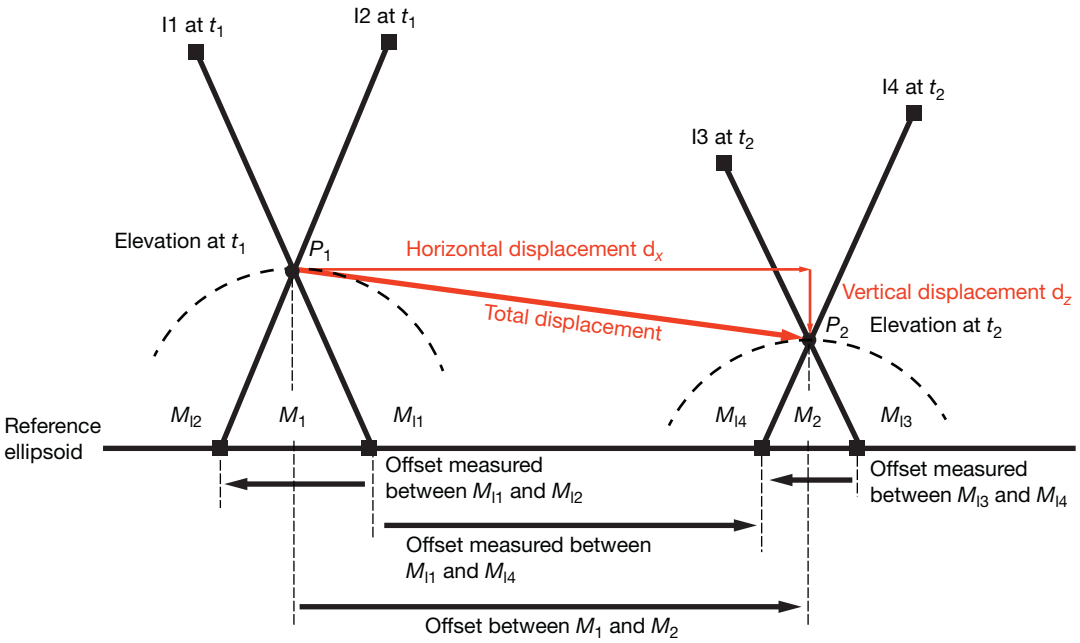


Figure 10 Given two pairs of stereo images (I1 and I2) and (I3 and I4), respectively, acquired at times t_1 and t_2, the 3-D displacement of a point P at the Earth's surface can be retrieved from the apparent offsets measured between each image pair projected onto a reference ellipsoid. Point P_1, which lies at the Earth's surface, is projected at M_1 and M_2. After deformation of the Earth's surface, P_1 is displaced to P_2, which will be projected at M_3 and M_4 from images I3 and I4. Knowing the position of the optical center of the imaging systems using the imaging system ancillary data, the 3-D position of P_1 and P_2 can be triangulated, from which the 3-D displacement vector from P_1 to P_2 can be deduced. The procedure also yields a determination of the elevation of point P at epochs t_1 and t_2. If the elevations at epochs t_1 and t_2 are known or assumed, then the horizontal displacement is directly determined from measuring the offset between the orthoprojections M_1 and M_2 of point P at epochs t_1 and t_2. In that case, only two images are needed.

inverted for the topography (one unknown) and the 3-D vector field (three unknowns). The problem is overdetermined because the topography should contribute to an offset along the epipolar direction at each pixel. Thus, the two components of the offset field should add redundant information on the topography. The procedure can be extended to the joint analysis of any n pairs of images following the bundle adjustment techniques described in the preceding text, which are customary in photogrammetry (Wolf and Dewitt, 2000). This is thus the most general and accurate approach with optical images. The horizontal displacement can also be determined from measuring the offset between the orthoprojections M_1 and M_2 of point P at the two epochs of image acquisition.

In the case of satellite images with near-vertical incidence, it is often assumed, provided that a reliable DEM is available as well, that the stereoscopic distortions due to topographic errors or to changes of elevation between the two epochs are negligible. In that case, only two images in addition to the DEM are needed. Only the horizontal displacement field can be determined. Vertical displacements cannot be determined in that case.

3.13.4.2 Algorithms for 3-D Matching

Various algorithms are available, which in principle can be used to match optical remote sensing datasets directly in 3-D. The 3-D surface matching problem is a well-covered topic in computer vision, computer graphics, and medical imaging (Besl and McKay, 1992; Grenander and Miller, 1998; Zhang, 1994). The goal is to determine a nonrigid spatial transformation that maps a surface onto another surface. The difficulties result from the fact that the problem is ill-posed in general and that the sampling 'grids' of the surfaces are in general independent (requiring some sort of interpolation). Because sampling does not satisfy the Nyquist conditions in general, as the real surface has always irregularities at a scale smaller than the GSD, the interpolation is always approximate.

The ill-posedness always requires some regularization strategy. This can be achieved by a priori assumptions on the transformation. For example, it might be assumed that the transformation is approximated locally by a rigid body translation and rotation. The matching problem is regularized if the scale at which this assumption is supposed to hold is significantly larger than the sampling distance and if the surface is nonplanar, neither cylindrical nor spherical, at this scale. Another more general strategy consists in defining a regularization energy penalizing large deformations, for example, by defining an 'elastic' energy so as to penalize bending and stretching of the surface during transformation (Grenander and Miller, 1998). Different algorithms are available, which might be adapted to geodetic optical remote sensing.

For example, the ICP algorithms (Besl and McKay, 1992) were successfully applied to reconstruct the displacement field of a slow-moving rockslide using terrestrial LiDAR data (Oppikofer et al., 2009; Teza et al., 2007). Nissen et al. (2012) had, for example, evaluated the performance of the ICP when applied to airborne LiDAR data. They found the

algorithm based on the point-to-plane metric of Chen and Medioni (1992) to perform best, so we briefly describe this particular algorithm.

For each point in the first dataset, the closest point in the second dataset is determined. For all points within a prescribed window, the rigid body transformation is determined so as to minimize the squared sum of the distances, l_i, between each point P_i of the second dataset and the tangential plane at its paired point M_i in the first dataset (**Figure 11**):

$$l_i = \vec{n_i} \cdot \overrightarrow{\phi(M_i)P_i} \quad [5]$$

The first iteration yields the rigid body transformation, ϕ_1, that minimizes the quantity $\sum l_i^2$. The process is iterated until some minimum is reached and the local transformation is the composition of the transformations determined at each iteration. This algorithm is very effective, converges better, and is less susceptible to yielding a local minimum than the original closest point metric (Low and Lastra, 2003).

Nissen et al. (2012) carried on synthetic tests in which they applied a known displacement to a subset of the B4 dataset, which was acquired along the major faults of the San Andreas Fault system in central and southern California (Bevis et al., 2005). The dataset consists of a point cloud with a sampling rate of about 2 points m^{-2} (mean GSD of 0.7 m) with nominal uncertainties of 25 cm on horizontal positions and 6 cm on elevation. Using sliding windows of $100 \times 100\ m^2$, within which the transformation is approximated by a single rigid body transformation, they were able to recover the imposed displacements with 1-σ uncertainties of 13 and 15 cm for E and N displacements and of 4 cm for vertical displacements. In this particular example, the uncertainty on horizontal displacement is estimated to be about 1/5 of the GSD and the uncertainty on vertical displacement, which is on the order of the nominal uncertainty on elevation measurements. This particular test was carried out by applying a known displacement field to the original LiDAR dataset. The advantage of this approach is that it does not require any explicit resampling of the original dataset. One major inconvenience is that there is no proof that the algorithm converges. The algorithm is very sensitive to noise because the determination of the normal vectors is very sensitive to horizontal registration errors and elevation errors. It can very easily get trapped in local minima especially when ground displacements or registration errors are in excess of the GSD. So even in the case of terrestrial LiDAR data, for which co-registration can be achieved with a better accuracy than with airborne data, ICP techniques do not perform better than the 2-step procedure involving first the determination of horizontal offsets from a 2-D matching algorithm (Daehne and Corsini, 2012).

3.13.4.3 Algorithms for 2-D Matching

This section presents an overview of different matching methods commonly used to measure the deformation between images of the same scene. The literature on image matching is abundant. The reader is referred, for example, to reviews by Zitova and Flusser (2003) or Scharstein and Szeliski (2002) (see also the associate website http://vision.middlebury.edu/stereo/). Here, we only mention those techniques that have already proven suitable to remote sensing applications for Earth sciences. In the context of this review, image matching is used to measure the disparity field that best morphs a slave onto a master image.

3.13.4.3.1 Homogeneous rotation and heterogeneous translation

Any transformation of a continuous field can be locally approximated by a homogenous transformation, that is, the combination of a rigid body rotation and a translation. In geodetic imaging from remote sensing data, it is generally admitted that the rotation component is homogeneous so that strain results dominantly from spatial variations of the

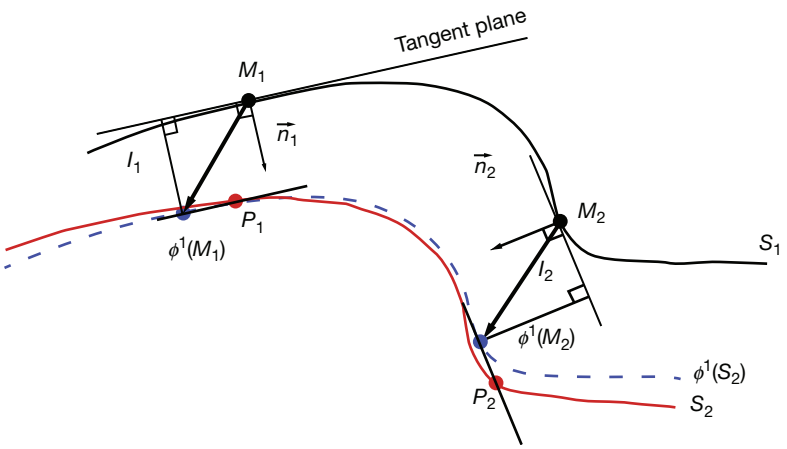


Figure 11 Matching of 3-D from the Iterative Closest Point technique, using the plane to point error metric (Low and Lastra, 2003). We consider that the surface has been sampled at two epochs t_1 and t_2. For each point M_i in the first dataset, the closest point P_i in second dataset is determined. For all points within a prescribed domain, the rigid body transformation ϕ is determined so as to minimize the squared sum of the distances, l_i, between each point P_i of the second dataset and the tangential plane at its paired point M_i in the first dataset (eqn [5]). So, the first iteration yields the rigid body transformation, ϕ_1, that minimizes the quantity $\sum l_i^2$. The process is then iterated until some minimum is reached. The local transformation is the composition of the transformations determined at each iteration.

translation component. As a result, the common practice is to first determine and correct for the homogeneous rotation component. This is generally achieved through the orthorectification procedure as, in most Earth sciences applications, local residual rotations are generally small. The next step is to determine the heterogeneous translation component. Any heterogeneous translation, provided that it is a continuous differentiable functional, can be locally approximated by a homogeneous translation. We therefore define matching as finding a globally nonrigid deformation between data, but that is approximated locally by a rigid translation. We will see that these assumptions and approximations hold for the applications reviewed here.

3.13.4.3.2 Optical flow

Optical flow methods were introduced by Horn and Schunck (1980) and Lucas and Kanade (1981), and many different implementations have been proposed since (e.g., Sun et al., 2010). The basic idea behind these methods is that the radiometric differences between the master and slave images are only due to plane deformation of the scene. This assumes that other factors of radiometric changes due to the imaging system and scene illumination have stayed unchanged or corrected for.

Let us consider the intensity i_1 (respectively i_2) measured a pixel location X,Y in master image 1 (respectively slave image 2). One may then write

$$i_2(X,Y) = i_1(X - d_X, Y - d_Y) \quad [6a]$$

where (d_X, d_Y) is the disparity vector field in the image space describing the heterogeneous translational transformation, which maps image 1 onto image 2. For small offsets, this equation can be approximated from its Taylor expansion to first order yielding

$$i_2(X,Y) - i_1(X,Y) \approx -d_X(X,Y)\frac{\partial i_1}{\partial X}(X,Y) - d_Y(X,Y)\frac{\partial i_1}{\partial Y}(X,Y) \quad [6b]$$

This equation shows that the deformation between two images is encoded in the image's brightness differences, just like eqn [4] that shows that horizontal advection of the ground surface is encoded in elevation changes.

This equation yields an ill-posed problem, as only the component of the offset vector field parallel to the image brightness gradient $(\partial i_1/\partial X(X,Y), \partial i_1/\partial Y(X,Y))$ can be determined. The problem can be regularized if solved at the scale of a local window, assuming that the disparity field is constant over a certain area (Lucas and Kanade, 1981) or using a global regularization approach (Horn and Schunck, 1980).

Under ideal conditions, the performance should only be limited by the radiometric noise. So, in theory, disparities might be measured with an accuracy better than 1/100 of the pixel size with 8-bit images and relatively small sliding window size (say 11×11 pixels) (Sabater et al., 2012). The technique has proven efficient and is adapted to measure strain from photogrammetry in laboratory analogue experiments (Bernard et al., 2007). This approach fails if disparities exceed about 1 pixel, as in this case, the Taylor approximation is not valid anymore. This problem occurs, for example, when the two images have different view angles and the surface roughness at the pixel scale is large (e.g., with high-resolution images of urban areas). This approximation also fails along a fault trace as the displacement field is locally discontinuous. The technique is very sensitive to variations of brightness not due to deformation of the scene. Optical flow methods have been extended to higher-order deformations, in particular, to the measurement of locally affine deformations and also to account for slight contrast variations (Broxton et al., 2009; Sabater et al., 2012). However, optical flow methods are often not robust enough to the strong illumination differences encountered in multitemporal remote sensing imagery, and they are therefore seldom used for geodetic imaging.

3.13.4.3.3 Statistical correlation

The idea behind statistical correlation is to use the Pearson's statistical correlation coefficient between an image patch taken in the master image and a multitude of neighboring candidate patches in the slave image (e.g., Barnea and Silverma, 1972). This technique is the basis for the particle imaging velocimetry method used to track fluid flow in fluid mechanics (Dudderar and Simpkins, 1977; Willert and Gharib, 1991) or sample deformation in experimental mechanical engineering (Hild and Roux, 2006). The technique has been used in Earth sciences, for example, to track glaciers, earthflows, and oceanic currents (Aryal et al., 2012; Debella-Gilo and Kaab, 2011; Marcello et al., 2008; Scambos et al., 1992).

The matching position of the two patches, hence the displacement between the patches, is found when the cross correlation attains its maximum. In order to take into account contrast and brightness variations, the cross correlation coefficient is normalized:

$$\rho(X,Y) = \frac{\sum_{X,Y}\left(i_2(X,Y) - \overline{i_2}\right)\left(i_1(X - d_X, Y - d_Y) - \overline{i_1}\right)}{\sqrt{\left[\sum_{X,Y}\left(i_2(X,Y) - \overline{i_2}\right)\right]^2\left[\sum_{X,Y}\left(i_1(X,Y) - \overline{i_1}\right)\right]^2}} \quad [7]$$

In practice, this simple formulation has been found to be one of the most robust against noise, affine changes of illumination, and temporal changes between images, and it is used in most implementations of image matching for remote sensing data (Marcello et al., 2008). There also exist a wide range of variations of correlation algorithms, with most variations depending on whether the correlation score is invariant by linear contrast changes and whether it uses an L1 or L2 norm (Zabih and Woodfill, 1994).

In a discrete correlation scheme, the slave correlation window can be seen as a moving window, moving with a step of 1 pixel at a time, therefore only sampling potential translation between master and slave images with integer displacements. Subpixel approximation is usually achieved by interpolating, or approximating, the maximum of the correlation peak with a quadratic or a Gaussian function (e.g., Debella-Gilo and Kaab, 2012). Although interpolation of the correlation maximum improves the correlation accuracy, it is, however, biased, and depending on the specific implementation, accuracy is often limited to about 1/5–1/4 of the pixel size. This bias can be

understood if we consider the correlation function, which involves the product of the master and slave images. Therefore, the correlation function exhibits a frequency support that is twice as large as the one of the images. Therefore, in order to avoid aliasing in the correlation function, one should in principle upsample both master and slave images by a factor of two. In practice, due to memory constraints, this upsampling is rarely implemented, leading to small biases. Another possible alternative to remove the subpixel bias is to iterate the correlation scheme with warping of the slave image between iterations. As it can be shown that the aliasing bias is always a fraction of the quantity to be measured, iterating the measurement will lead to a negligible bias. In practice, one iteration removes enough bias to allow measurement accuracy better than 1/10 of the pixel size but has an increased computational cost and assuming the warping does not introduce additional artifacts.

3.13.4.3.4 Phase correlation

Phase correlation methods take advantage of the Fourier shift property, whereby a translation in the image domain is equivalent to a phase shift in the Fourier domain.

Accordingly, the Fourier transform of eqn [6a] yields

$$\widetilde{i}_2(\omega_X, \omega_X) = \widetilde{i}_1(\omega_X, \omega_X)\mathrm{e}^{-j(d_X\cdot\omega_X + d_Y\cdot\omega_Y)} \qquad [8]$$

The local translation (d_X, d_Y) can therefore be retrieved using the inverse Fourier transform (noted $\mathbb{F}^{-1}$) of the images cross spectrum, such that

$$\delta(X-d_X, Y-d_Y) = \mathbb{F}^{-1}\left(\frac{\widetilde{i}_1(\omega_X, \omega_X)\widetilde{i}_1^*(\omega_X, \omega_X)}{\left|\widetilde{i}_1(\omega_X, \omega_X)\widetilde{i}_1^*(\omega_X, \omega_X)\right|}\right) \qquad [9]$$

When the displacement (d_X, d_Y) is not an integer number of pixels, an interpolation problem arises, and proper peak interpolation often requires iterating the correlation with image resampling (Vadon and Massonnet, 2000).

Another solution to avoid interpolation problems is to directly solve for the displacement dx in the Fourier domain via an inverse problem. Indeed, one can solve for (d_X, d_Y) that minimizes

$$\int\left[\frac{\widetilde{i}_1(\omega_X, \omega_X)\widetilde{i}_2^*(\omega_X, \omega_X)}{\left|\widetilde{i}_1(\omega_X, \omega_X)\widetilde{i}_2^*(\omega_X, \omega_X)\right|} - \mathrm{e}^{j(d_X\cdot\omega_X + d_Y\cdot\omega_Y)}\right]^2 \mathrm{d}\omega \qquad [10]$$

This nonlinear minimization problem can be efficiently solved using linearization if initialized close to the solution given by the inverse Fourier transform from the preceding text.

Comparing with a statistical correlation, phase correlation methods are usually computationally more efficient taking advantage of the FFT algorithm, and since no aliasing problem occurs in this formulation, it has the potential to be highly accurate, often providing results with accuracy on the order of 1/20–1/10 of the pixel size using small window sizes (e.g., 32×32 pixels). The normalization in the Fourier domain has also shown to be very robust against illumination changes and even against sharp contrast differences. Overall, practice has shown that phase correlation has the potential to be more accurate and less sensitive to contrast or shadow changes than statistical methods. They, however, tend to be more sensitive to noise, and correlation windows need to be larger than 16×16 pixels, which reduces the spatial resolution of the displacement field retrieved. This approach has proven efficient to measure coseismic deformation from satellite optical images, glacier flows, and earthflows (Leprince et al., 2007, 2008a; Van Puymbroeck et al., 2000).

3.13.4.3.5 Regularized solutions and large displacements

Particularly in the context of topography extraction or in the context of 3-D measurement of displacement fields, large pixel offsets may need to be measured. To lower the complexity of the matching algorithm by reducing its search space, it is customary to proceed in a multiscale fashion, where master and slave images are downsampled by a factor that allows a reasonable complexity for the matching algorithm (e.g., Pierrot-Deseilligny and Paparoditis, 2006). The offset field found at the coarser scales is then upsampled to higher scales, where the slave image at higher scales is warped according to the offset field measured at lower scales. Iteratively, only a differential offset field needs to be computed at each scale, lowering the complexity of the algorithm.

This multiscale approach also allows matching deformations that can depart significantly from a local translation. Indeed in this context, the condition to be met for good matching is only that the deformation field between successive scales be locally approximated by a translation.

One major drawback of multiscale schemes is that errors can easily be propagated between scales. The matching algorithm must therefore be augmented with a regularization term to ensure that spurious matches do not occur and to ensure that every point at every scale is assigned a likely match, to be propagated. As a result, missing matches cannot be tolerated within a multiscale approach.

The state-of-the-art regularization that is the most widely used at the time this paper is being written is the regularization on the L1 norm on the gradient of the offset field, solved by Semi-Global Matching (Hirschmuller, 2005; Hirschmuller and Scharstein, 2009). This approach offers a good compromise between the maximization of the correlation coefficient and the smoothness of the offset field. Such regularization technique is usually applied to image-based matching functions such as the normalized cross correlation presented in the preceding text (Pierrot-Deseilligny and Paparoditis, 2006), entropy, or census matching.

3.13.5 Geometric Modeling and Processing of Passive Optical Images

3.13.5.1 The Orthorectification

The orthorectification is the process that projects an image on the topography surface by assigning absolute geolocation coordinates (x,y,z) to each image pixel. An orthorectified image, or orthoimage, is therefore free of stereoscopic effects, since it simulates an image as if each pixel had been acquired with a viewing angle normal to the projection datum. To achieve this result, the orthorectification is composed of two processing

steps: (1) computing the mapping between the image pixel coordinates and the ground coordinates and (2) resampling of the image according to this projection mapping.

3.13.5.1.1 The orthorectification mapping

Generating the orthorectification mapping is solving a ray-tracing problem, which is a geometric problem. It requires the knowledge of the camera geometry so that the direction of light rays hitting every pixel in the image can be established. Given a camera model, its position, and its orientation in space, it is then possible to determine which points on the ground reflected the sunlight that hits a particular pixel using the reverse light propagation principle. Simple models usually assume a standard pinhole camera model with light propagating along straight lines in the atmosphere. More complex models will include the camera optical and sensor distortions, the variation of the atmospheric refraction index, and the relative speed of the camera with the speed of light to determine precise orthorectification mappings. For instance, a simple model for a push broom sensor (Figure 9) can be given by the following equation (Leprince et al., 2007):

$$M(p) = O(t) + \lambda \cdot T(t) \cdot R(t) \cdot \vec{u}\,(p) \qquad [11]$$

where $M(p)$ is the ground point seen by the pixel of coordinates $p(X,Y)$, t is the time at which the pixel p was acquired, O is the location of the optical center when the pixel p was acquired, $\vec{u}\,(p)$ is the reverse direction of the light ray for the pixel p, $R(t)$ is the 3-D rotation matrix recording the 3-D rotation of the camera in space at time t, $T(t)$ is the system reference change matrix from the camera (orbital) to the terrestrial reference system, and λ is the distance between the optical center and the object seen by the pixel p.

Parameters describing the camera, that is, the set of vectors u, are defining the internal orientation (IO) model of the system, and the parameters describing the camera positions and orientations in space, that is, O and R, are defining the external orientation (EO) model of the system.

Knowledge of a fine model of the topography is also needed to determine λ, as M lies at the intersection between the imaging pointing vector $T \times R \times u$ and the topographic surface.

3.13.5.1.2 Resampling the image

Producing an orthoimage implies producing an image that is regularly sampled in a given georeferenced system, so it can be displayed. However, because of the topography variations and changes in camera viewing angle, the mapping associating the pixel coordinates to the ground coordinates is often highly irregular (Figure 12). To avoid solving a complex irregular resampling problem, it is often convenient to assume that the sensor delivers a regularly sampled image. This assumption can be considered exact for frame cameras as the sensor is composed of a unique and flat focal plane, but it is only a local approximation for push broom sensors that exhibit varying attitude over time, with the quality of this assumption depending on the attitude stability of the sensor in time. Therefore, instead of attributing ground coordinates to each image pixel, we preferentially solve the inverse problem, which attributes pixel coordinates to every point of the orthorectification grid (Leprince et al., 2007).

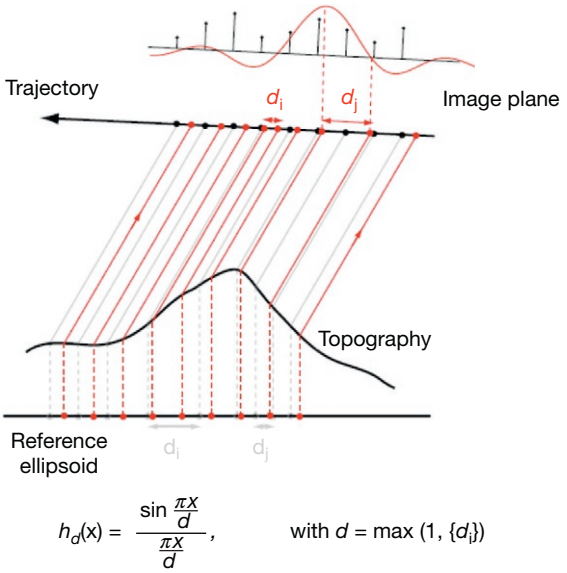


Figure 12 Due to the topography, the orthoprojection of the ground spots sensed by regularly spaced CCD is irregularly spaced. Instead of attributing ground coordinates to each image pixel, we solve the inverse problem, which attributes pixel coordinates to every point of the orthorectification grid (Leprince et al., 2007). The intensity at those points can be estimated with well-known resampling kernels (ideally a sinc function) in the image space.

This method allows the use of more traditional, and much simpler, resampling methods with well-known resampling kernels. When the sampling density of the orthoimage is similar, or less than, the raw acquisition, the image resampling simply turns to an interpolation problem, where the raw image simply needs to be interpolated at the pixel coordinates given by the orthorectification mapping. One needs to be careful when the sampling density of the orthoimage is less than the sampling density of the raw image. In this case, the resampling kernel can be approximated by the interpolation kernel, dilated by the change in sampling density (Figure 13). Intuitively, the change in the sampling density can be seen as the local ratio between the raw image sampling and the orthogrid sampling. This operation can be seen as a local aggregate of the raw image pixels whenever the orthogrid resolution is coarser than the raw image resolution. Formally, the local dilation of the resampling kernel can be represented as the Jacobian of the orthorectification mapping, and the sampling density variation is represented by the absolute value of its determinant (Leprince et al., 2010).

How to adapt the resampling kernel to the orthomapping applies to any resampling kernels. In practice, the choice of a particular resampling kernel will depend on the constraints of the application at hand, with typically a compromise between accuracy and processing time (or algorithmic complexity). For geodetic applications, it is recommended to select a resampling kernel belonging to the cardinal-sine family. The cardinal-sine function (sinc) is the theoretically exact sampling kernel for regularly spaced samples.

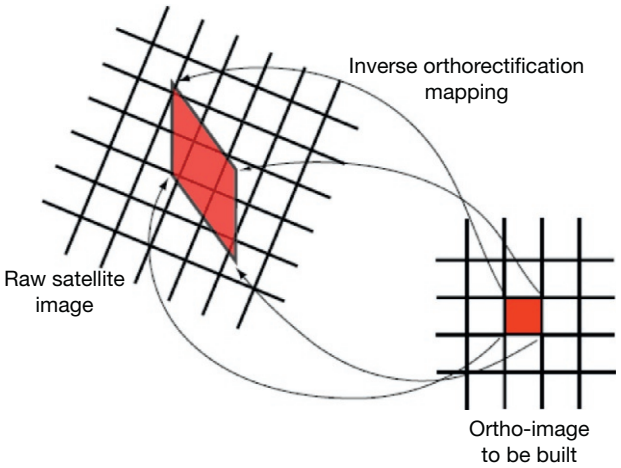


Figure 13 The resampling kernel used to calculate the orthoimage needs to be adjusted to the distortions induced by the topography. Formally, the local dilation of the resampling kernel can be represented as the Jacobian of the orthorectification mapping, and the sampling density variation is represented by the absolute value of its determinant. Leprince S, et al. (2010) Rigorous adaptive resampling for high resolution image warping. In: *IGARSS: Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1943–1946.

$$h_d(X) = \frac{\sin\frac{\pi x}{d}}{\frac{\pi x}{d}}, \qquad \text{with } d = \max(1, \{d_l\})$$

It exhibits low phase distortion and therefore introduces less subpixel biases in the orthoimage (Van Puymbroeck et al., 2000). Whenever the geometric accuracy of the orthoimage is important, truncated or weighted sinc resampling kernels are usually recommended.

3.13.5.2 Bundle Adjustment

The orthorectification mapping relies on the precise knowledge of the parameters defining the interior and exterior orientations of the optical system. Interior orientation parameters are usually determined through optical bench calibration, and EO parameters are usually recorded during the image acquisition using GPS and IMU devices. Unfortunately, these parameters (in particular for the exterior orientation) are often not known with enough accuracy to provide geodetic measurements from optical images. They are therefore usually refined using GCP. GCPs are recorded during ground surveys, and they associate ground coordinates to recognizable features in the image. External parameters are then optimized such that, for a given pixel coordinates in a GCP, the distance between the ground coordinates predicted by the orthomapping for this given pixel and the actual GCP ground coordinate is minimum.

When presented with a system composed of frame cameras, the positions in space and orientations of the cameras are optimized, leading to six external parameters to be optimized per image. When the system is composed of scanning devices such as push broom sensors, the exterior orientations are composed of varying positions and orientation angles in time. In this case, it is customary to model the error on the positions and orientation angles as second-order polynomials varying in time, leading to 18 EO parameters to be optimized per image.

When several images of the same area are available, corresponding pairs of pixels between images, called tie points, can be selected. This selection is usually automatic using feature matching algorithms (e.g., SIFT and Forstner operator) (Forstner and Gulich, 1987; Lowe, 2004), and false matches are rejected using robust statistics (RANSAC) (Beckouche et al., 2011; Fischler and Bolles, 1981). However, visual verification of tie points is still customary in practical applications.

Here again, the viewing parameters of the imaging systems can be optimized thanks to the tie points. Formally, the objective is to find the sets of EO parameters such that the rays issued from the tie points intersect as closely as possible. However, in the absence of absolute GCPs, the solution of this system will not be well constrained and the optimization of the parameters is only relative. For instance, it is easy to see that in the absence of absolute GCPs, the optimization of the tie points could lead to a system that could be arbitrarily translated and scaled in space. Therefore, in the absence of GCPs, additional a priori information on the maximum deviation of the given exterior parameters is required. Whenever possible, it is usually best to jointly optimize a set of tie points and GCPs to insure precise absolute location of all images. This global optimization that consists of sets of images is called the bundle adjustment. This is a standard approach in photogrammetry when DEMs are produced from various sets of overlapping images (Wolf and Dewitt, 2000).

In our experience, performing a bundle adjustment is always necessary to deliver absolute geodetic measurements with an optical imaging system. The accuracy of the rectified imagery depends on the accuracy of the bundle adjustment, which itself depends not only on the accuracy of the tie points and GCPs provided but also on the quality of the sensor modeling. From a user's perspective, providing tie points and GCPs with subpixel accuracy (with accuracy better than the image pixel size) is always necessary to ensure a near-geodetic quality of the measurements.

3.13.5.3 Stereo Imaging

Stereoscopic imaging refers to the possibility of extracting the 3-D shape of a surface imaged using at least two images of the same scene from different viewpoints. A bundle adjustment of the images is often required to provide digital surface models with acceptable accuracy. Assuming that light travels along straight lines, the concept of stereo imaging can be described as solving a simple geometric problem (**Figure 9**). From the camera interior orientation, we know how to model the light rays hitting each pixel of the image. From the exterior orientation, we know the location in space of the optical centers and orientation of the cameras. Now, if we add another function to the system that provides matching pairs of pixels between the different images, it is possible to find the intersection of light rays between the corresponding pixels of the different images. The intersection of light rays therefore defines the location of the surface that reflected the light to the sensor. Repeated over the common pixels between the set of images, it becomes possible to reconstruct the imaged terrain in three dimensions.

It is clear that the higher the angle between intersecting rays (referred to as the stereoscopic, convergence, or parallax angle),

the better constrained the location of the intersection will be. However, large convergence angles are also likely to produce occlusions, where parts of the scene will be hidden behind high slopes. In addition, the matching algorithms tend to have lower failure rates when images are acquired with lower convergence angles, which minimize the relative distortion between images. In practice, it is preferable to keep the convergence angle between 10° and 45°, with the lower angles being used in areas with the highest topographic slopes to limit occlusions.

In stereo imaging, it is assumed that stereo images are acquired simultaneously or at least that potential changes between acquisitions can be neglected. In a multitemporal stereo set up, we add another step, where we can associate pixels from images acquired at different times. From a stereo pair at a given time, we determine the 3-D coordinates of the surface, we determine where this point on the surface is in the other stereo pairs, and the other stereo pairs are used to give the 3-D coordinates of the given point, at another time, hence producing a 3-D displacement vector (Figure 10).

3.13.5.4 Processing Flowchart

When one uses only two images bracketing a deformation event, only the horizontal displacement field can be recovered. If the two images were taken from exactly the same viewpoint, displacement of the ground parallel to the image plane (i.e., horizontal) would be the only cause of distortion in principle. The topography is, however, needed in order to register the measured displacement through orthorectification of the two images.

This is never the case in practice. At best, the two images have close view angles but the look angle vectors are variable so that stereoscopic effects are always present. In principle, one would need to know the topography at the two epochs. The offsets measured between the two orthorectified images would then represent horizontal displacement in the geographic coordinate system defined by the GCP used for registration. In case of two images with close and near-vertical incidence view angles, the distortions are small, and it is then valid to use even a coarse DEM and neglect topographic changes. As we will see in the examples later, this assumption is often challenged. Images acquired within a few degrees of nadir looking help mitigate this constraint.

Within this setup, the analysis of the images can be performed either in the ground geometry or in the image geometry. Performed in the image geometry, it would mean that the slave image would be projected in the master image geometry using the DEM to suppress stereoscopic parallax, and the disparity map, in the master geometry, would have to be orthorectified according to the master geometry in order to produce a ground displacement map.

Performed in the ground geometry, the master and slave images are both orthorectified, and the disparity map is computed between the orthomaster and the orthoslave images. The main advantage is that the disparity map therefore directly produces the ground displacement map, without the need to further resample the disparity.

We generally prefer using the ground geometry, as is the case in the examples reviewed in the succeeding text, because the implementation is less dependent on the modeling of the sensors, making it easier to mix different types of imagery together; it also makes the different blocks of the processing chain more independent, allowing the use of potentially different software solutions along the processing chain; and it also avoids the delicate task of resampling the disparity field. However, the reader should be aware that disparity fields computed in the ground geometry may be slightly noisier than disparity fields computed in image geometry, because images are subject to less manipulation when left in image geometry.

The processing in ground geometry typically involves the following steps (see flowchart of Figure 14):

- Select tie points and GCP between the master image and the topography model.

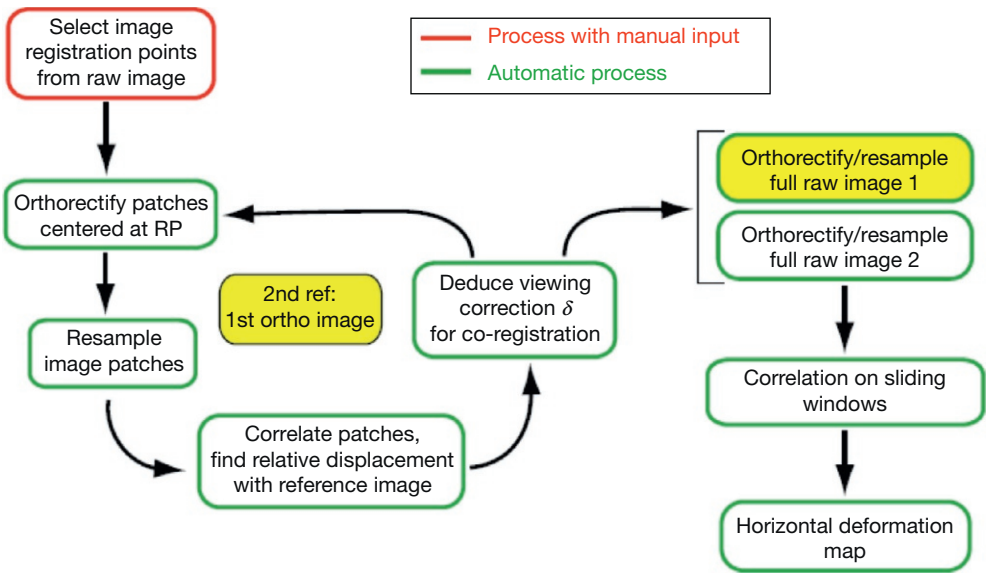

Select image registration points from raw image
Process with manual input
Automatic process
Orthorectify patches centered at RP
Orthorectify/resample full raw image 1
Orthorectify/resample full raw image 2
2nd ref: 1st ortho image
Deduce viewing correction δ for co-registration
Resample image patches
Correlation on sliding windows
Correlate patches, find relative displacement with reference image
Horizontal deformation map

Figure 14 Flowchart for the measurements of horizontal displacements from two optical images and a DEM. See text for details.

- Optimize the exterior orientation parameters of the master image according to the GCP selected and orthorectify the master image.
- Select tie points and GCP between the slave image and the orthomaster image, in areas of stable ground.
- Optimize the exterior orientation parameters of the slave image according to the GCP selected and orthorectify the slave image.
- Perform dense image matching between the orthomaster and orthoslave images to produce the displacement field.

When measuring 3-D displacements, it is assumed, by definition, that the DEM available is not sufficient to describe the ground motion. In this case, if a DEM is available, it will be used in the process as a seed to speed up computation by reducing the search range of the stereo processing.

As for the estimation of 2-D displacement fields, the estimation of 3-D displacement fields can be formulated in image geometry or in ground geometry. However, even in ground geometry, the disparity field recovered may need to be resampled because the large incidence angles commonly used for 3-D processing may not allow the derivation of a proper orthomaster image.

A typical processing chain for the estimation of 3-D ground motion can be the following:

- Select tie points between all stereo pairs.
- Select tie points between the multitemporal pairs in areas of stable ground only.
- Perform a bundle adjustment of all data.
- Orthorectify all images on a seed DEM.
- Perform image matching between orthoimages for stereo pairs and for multitemporal pairs.
- Triangulation of stereo pairs produces 3-D points.
- Matching of the multitemporal pairs associates 3-D points before and after the event, thereby producing 3-D ground displacement vectors.

3.13.5.5 Performance, Artifacts, and Limitations

Most artifacts and limitations to the measurement of ground deformation from optical imagery can be grouped in a small number of categories, which are geometric errors, topographic errors, improper placement of tie points/GCP, shadowing effects, and poor image matching. We point to the figures illustrating these various effects, which are commented later on in the text (chronological ordering of figures is not respected here). Quantitative estimates of these errors are provided when these examples are reviewed.

- Geometric errors

  Uncertainties in the geometry of the imaging sensor will introduce biases in the ground deformation measured. Common sources are
  - jitter/attitude artifacts due to push broom motion (unmodeled platform attitude variations) (Figures 21 and 22),
  - CCD misalignment or miscalibration in the focal plane (Figure 19),
  - scanning artifacts (when using film archives) (Figure 20(b)),
  - unknown film distortions (when using film archives) (Figure 20(a)).
- Topographic errors

  If only two images bracketing an event are available, then it is assumed that the DEM should properly account for the height modeling. If the DEM resolution is too coarse with respect to the image resolution, a bias will be introduced in the ground deformation measurement, which depends on the difference between the incidence angles of the two images (Figure 15).

  In general, topographic errors occur when as follows:
  - The DEM has insufficient horizontal or vertical resolution (Figure 17).
  - The topography between acquisitions has changed and a single DEM is used (Figure 28).
  - Images are not well registered to the DEM.
- Location of GCPs

  All the processes require optimization of the viewing geometry of the sensor (a least exterior orientation via bundle adjustment), and optimization requires GCP and tie points between images. One has to be careful to select these points on stable ground, or the bundle will try to absorb the ground deformation into the sensor modeling. Particular difficulties can arise when the swath of the imaging sensor is limited (e.g., airborne sensors) and when

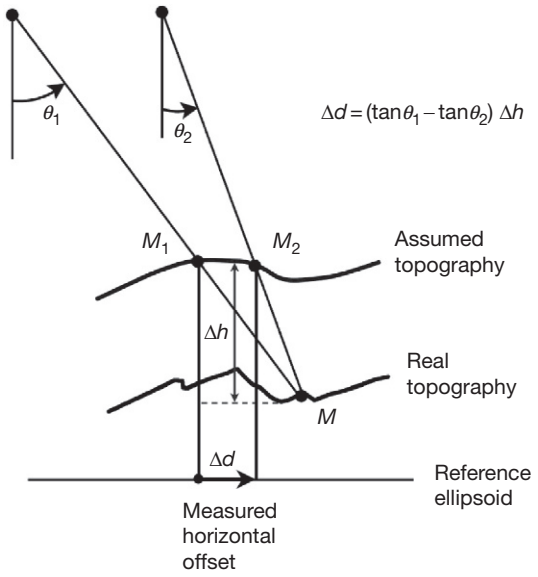


Figure 15 Sketch illustrating artifacts due to topographic errors in the case where surface displacements are measured from matching optical images orthorectified based on a preexisting DEM. Any DEM is a first-order approximation of the real topography. It follows that the position in the images 1 and 2 of a particular point M on the ground is determined by the real topography and the look angle vectors. If the view angles, θ_1 and θ_2, are different, the orthorectification procedure will yield positions M_1 and M_2, which are biased due to the difference between the real and assumed topographies. An apparent horizontal displacement will result. This bias will in particular be important with images with high incidence view angles, when the topography is rugged. The topographic error can be due to the uncertainty on the determination of the topography and to the advective transport of the topography.

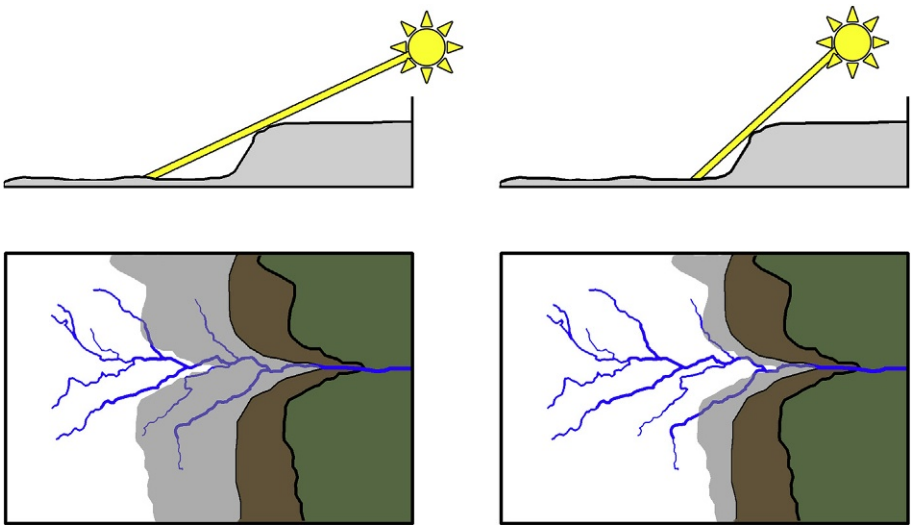


Figure 16 Sketch illustrating artifacts due to shadows. Generally, two nonsynchronous images of a same scene will show different shadows due to the difference of sun azimuth and sun elevation. Strong cast shadows may overwhelm ground features. If the texture of the image at the scale of the sliding window used for 2-D matching is dominated by such topographic shadows, then the image matching algorithm will likely track the moving shadows rather than the moving ground.

ground deformation is present in most of the field of view. When tie points/GCP is taken on moving grounds, long-wavelength biases are usually introduced in the deformation maps (Figure 25).

- Shadowing effects

  Although most Earth-observing satellites are placed on Sun-synchronous orbits (the satellite always sees a given point at the same local time), seasonal variations still induce changes in the orientation and length of shadows during the year. Problems may arise when strong cast shadows are present, as their contrast may overwhelm the contrast of the ground objects (Figure 16). If the texture of the image at the scale of the sliding window is dominated by topographic shadows, then the image matching algorithm will likely track the moving shadows rather than the moving ground, thereby introducing a bias (Figure 17). If using a matching algorithm with contrast invariance features, diffused shadows are usually not a problem.

- Image matching failure

  Whenever image matching fails, ground displacement cannot be estimated. Image matching can fail for a variety of reasons, most commonly because the scene has drastically changed between acquisition dates and corresponding points cannot be identified between images. Drastic changes are usually due to changes in snow/cloud/vegetation cover. Areas with steep slopes that create hidden parts (occlusions) also cannot be matched. The likelihood of occlusion increases with the increase of the off-nadir angle of the images. Occlusion problems are common in urban and mountainous areas.

As an illustration, Figure 17 shows the result of a 'blank' test (Leprince et al., 2007). Two SPOT 5 images from the same scene in the Mojave Desert area were chosen. The two images were acquired 6 months apart. No significant deformation event is known to have occurred in that period so that the measurements should only reflect the various sources of noise and bias mentioned in the preceding text. These images have a GSD of 5 m, incidence view angle of $-1.642°$ and $-1.726°$, and sun elevation of 68° and 33° and about the same sun azimuth. The two images were orthorectified using the 1/3 arcsec NED DEM from USGS, co-registered, and correlated following the flowchart of Figure 14. Offsets were calculated using the phase method using a 32×32-pixel sliding window with a step of eight pixels (80 m on the ground). The exterior orientation is satisfyingly modeled as we do not see the typical oscillating pattern due to jitter errors (roll, pitch, or yaw variation residuals). Offsets are at places correlated with the topography. They are too large to be explained by parallax effects given the near-nadir incidence. The artifacts, which reach up to a few meters, are seen only in the N–S component and are therefore more consistent with a shadowing effect. No other biases are visible, meaning that precise orthorectification and co-registration have been achieved. Other sources of errors result in some sort of a white noise with no spatial pattern. The histogram of offsets measured over the whole scene yields a close to normal distribution with a mean of 5.8 cm and a variance of 80 cm (Figure 18). The mean is 5 cm and the variance is 57 cm, on the E–W component that is less affected by shadowing. These errors are about one order of magnitude larger than those expected from the performance of the matching technique and are probably mostly due to natural changes of the scene.

Figure 20 illustrates the noise and artifacts of geodetic imaging with air photos. These images have a ground resolution of the order of 50 cm. The blank tests presented in this figure show that film distortions generally prevent measuring deformation at a wavelength larger than about 1 km. At shorter wavelengths, the noise level yields an rms of only 5 cm (so about 1/10 of the pixel size) essentially due to scan artifacts, temporal decorrelation, and shadow effect. This performance

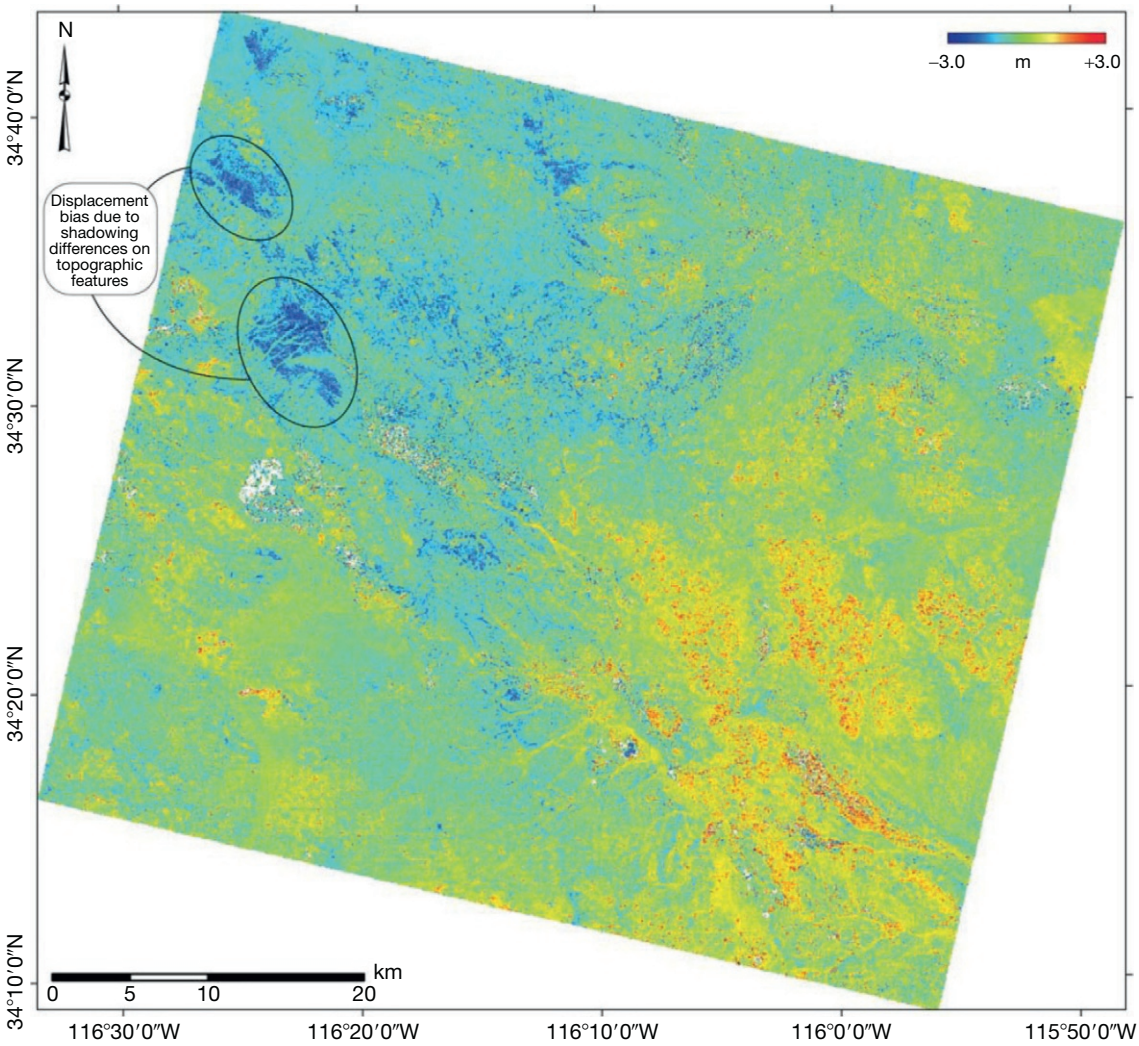


Figure 17 N–S component of offsets between two orthorectified SPOT 5 images measured from the phase correlation method using a 32×32-pixel matching window (Leprince et al., 2007). The images show the 1999 Hector Mine earthquake area in the Mojave Desert. They were taken 6 months apart on 26 July 2002 and 24 January 2003. No significant earthquake occurred over that period and postseismic displacements are negligible. Shadowing biases are mostly visible in this component since the Sun azimuth of the two images is mostly N–S-oriented (127.72° and 158.15°). Decorrelation points are shown in white.

allows detecting and measuring surface slip on faults with as little as 10–20 cm of displacement (Ayoub et al., 2008; Michel and Avouac, 2006). These performances are achievable only if the quality of the scan is sufficient. **Figure 19(b)** shows the typical pattern and amplitude of scan artifacts resulting from a lower-quality scanning (here, the slave image has a nominal ground resolution of 80 cm). In that case, scan artifacts reach up to 20 cm and are clearly the main source of errors at short wavelengths (<1 km).

3.13.6 Applications to Coseismic Deformation

3.13.6.1 Usefulness of Coseismic Deformation Measurement from Image Geodesy

Earthquakes are associated with sudden slip events on faults. Large earthquakes therefore produce permanent surface deformations, which might be measured from geodesy or remote sensing. When a fault rupture during an earthquake reaches the surface, as is often the case for Mw > 7 crustal earthquakes, surface fault slip might also be measured directly from field investigations (e.g., Sieh et al., 1993). Field investigations take time so that this information is generally not taken into account in early source model determination. They are known to provide key insight regarding faults geometry and the distribution of slip and help constrain earthquake source models, which are often ill-constrained without this information (e.g., Konca et al., 2010). Detailed observation of recent ruptures is also important as it provides clues for the interpretation of past ruptures as revealed from paleoseismic investigations (Yeats et al., 1996). In the field, fault slip is generally estimated from the offset of linear features assumed initially continuous across the fault such as a road, a fence, or a terrace riser. Such features are not densely distributed so that measurements are often sparse, typically a few per kilometer. In

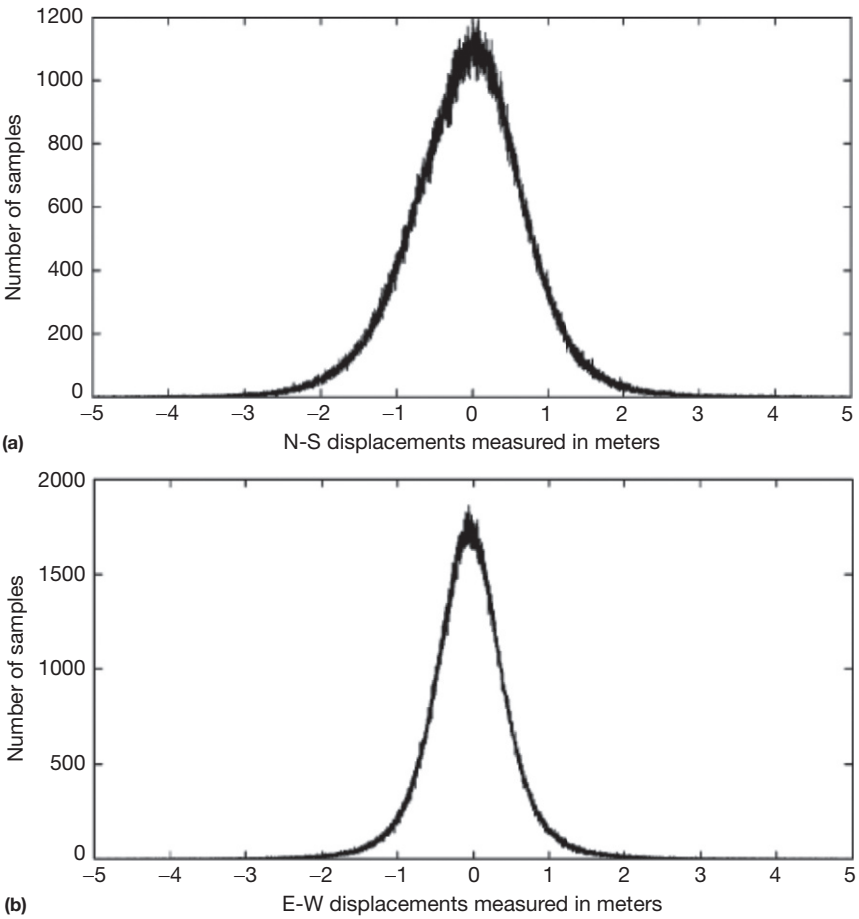


Figure 18 Histograms of the offsets shown in **Figure 17** (Leprince et al., 2007). (a) N–S component ($\mu(\Delta NS) = -0.058$ m and $\sigma(\Delta NS) = 0.80$ m). (b) E–W component ($\mu(\Delta EW) = -0.051$ m, $\sigma(\Delta EW) = 0.57$ m). The histograms are nearly Gaussian and can be seen as reflecting the noise on the measurements. The noise due to the correlation technique is probably only a very minor factor. Given the close incidence angles of the two images, topographic bias can only account for at most a few centimeters. The large difference in the Sun elevation then largely contributes in biasing the correlation measurements on topographic features.

addition, while the vertical and strike-slip component of fault slip can often be determined with some accuracy, the component perpendicular to the fault trace is most generally not measurable. Finally, these measurements are possible only across well-localized faults. If anelastic deformation is distributed, it is generally not possible to make any reliable measurement, unless the initial geometry of the linear feature can be assumed with confidence (e.g., in the case of well-aligned electrical poles). The measurement of surface fault slip from correlating images acquired before and after an earthquake is advantageous in that it does not suffer from any of those limitations and could in principle be produced quite early after an earthquake and taken into account for damage assessment. As it provides surface displacement in map view, the technique can in principle be self-sufficient to determine the distribution of fault slip at depth. This can in principle be achieved based on the theory of dislocations in an elastic half space (Okada, 1985) as commonly done with geodetic measurements and SAR interferograms. Geodetic imaging from optical methods is very complementary to these techniques as it provides information on near-fault deformation where geodetic measurements are rarely available and where SAR interferometry generally fails due to decorrelation induced by ground deformation or damages or due to strain exceeding the fringe rate limit of one fringe per pixel (**Chapter 3.12**; Michel et al., 1999).

Numerous applications based on passive optical images have now been published (e.g., Avouac et al., 2006; Ayoub et al., 2009; Binet and Bollinger, 2005; Chini et al., 2011; Copley et al., 2011, 2012; Dominguez et al., 2003; Karabacak et al., 2011; Klinger et al., 2006; Konca et al., 2010; Michel and Avouac, 2002; 2006; Oskin et al., 2012; Taylor et al., 2008; Van Puymbroeck et al., 2000; Wei et al., 2011). A few major faults in California and elsewhere have been surveyed with LiDAR (Bevis et al., 2005; Phillips et al., 2008; Prentice et al., 2003). These data have been exploited to analyze the geomorphic signature of past earthquakes (Arrowsmith and Zielke, 2009; Hudnut et al., 2002; Zielke et al., 2010, 2012) or for method development (Borsa and Minster, 2012; Nissen et al., 2012). To date, the El Mayor–Cucapah earthquake of 2010 is the only earthquake for which LiDAR data had been acquired before the event (Oskin et al., 2012). We review in the succeeding text a selection of examples to illustrate the performance and limitations of the optical geodetic approach to measure coseismic deformation.

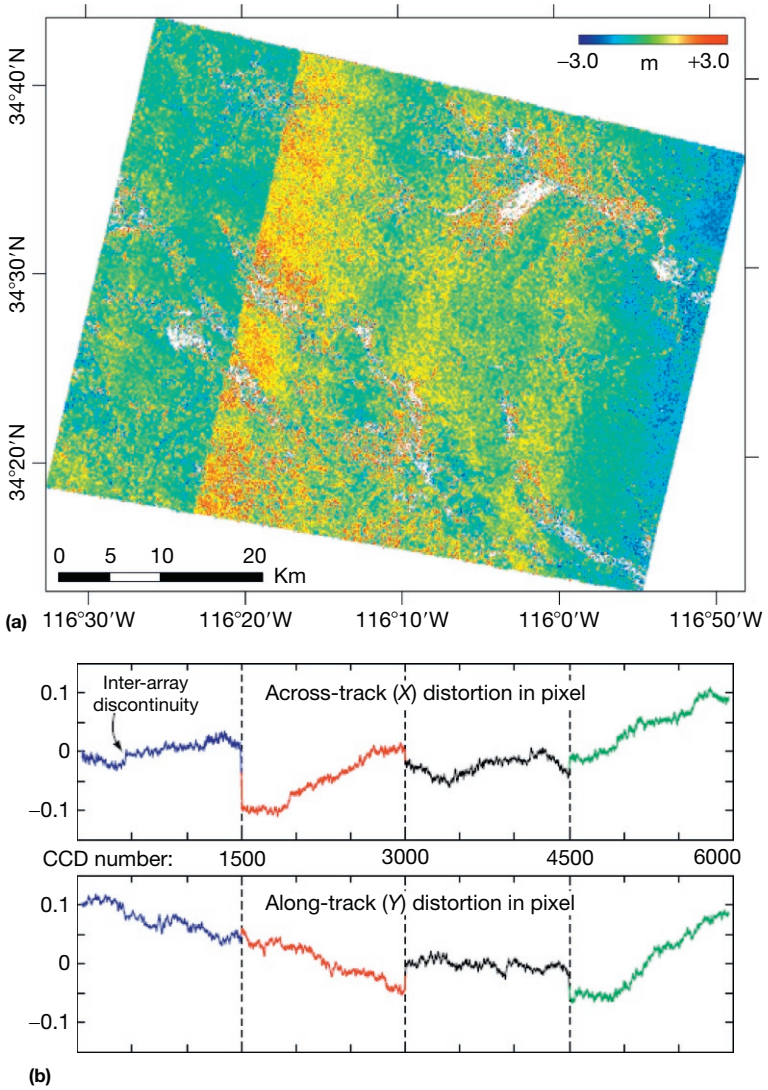


Figure 19 (a) E–W component of the offset field measured from the subpixel correlation of a reference SPOT 5-HRG1 5 m panchromatic image acquired on 24 January 2003 and a slave SPOT 4-HRV1 10 m panchromatic image acquired on 11 March 2000 of same area of Mojave Desert as in **Figures 22** and **19** (Leprince et al., 2008b). Both images were orthorectified using the 1/3 arcsec NED DEM from USGS. Thirty subpixel GCPs were used to tie the SPOT 5 image to the DEM, and six subpixel GCPs were used to tie both orthorectified images together. Correlation analysis was performed on 32×32-pixel matching windows, sliding with a step of 8 pixels (80 m on the ground). Linear artifacts reveal the SPOT 4 CCD distortions. (b) Measured distortions, in pixels, in the focal plane of the SPOT 4-HRV1 panchromatic sensor derived from the offset field shown in Figure 16 (Leprince et al., 2008b). The CCD line sensor is composed of four CCD line arrays of 1500 pixels each (vertical dotted lines). Both across-track and along-track distortions are measured with an uncertainty below 0.01 pixel rms. Distortions of up to 0.12 pixel ($\sim$1.2 m on the ground) are estimated allowing for precise geometric calibration of the whole CCD line sensor.

3.13.6.2 Surface Displacement in 2-D due to the 1999 Mw 7.6 Chichi Earthquake, Measured from SPOT Images

As a first illustration, **Figure 21** shows the east–west (E–W) offset field measured from correlating two panchromatic SPOT images (pixel size of 10 m) acquired 29 January 1999 and 23 November 1999, respectively, 8 months before and 2 months after the earthquake (Dominguez et al., 2003). This earthquake resulted from thrusting on the Chelungpu fault along the western foothills of the Central Range in Taiwan. Field investigations provided tight constraints on the cartography of surface ruptures and the vertical and strike-slip component of fault slip across the fault trace (Chen et al., 2001). The component of the slip perpendicular to the fault strike could not be measured in the field (hence, the E–W component as the fault is striking approximately N–S).

This example is interesting in that a quite dense dataset of near-field GPS data was available (Yu et al., 2001). Also, it illustrates well the benefit of the technique and the importance of a good modeling of image geometry when GPS data are not available. Given that no significant preseismic deformation was observed from the GPS permanent stations and that only a few centimeters of horizontal postseismic displacements were measured in the near-field areas over the 3 months that

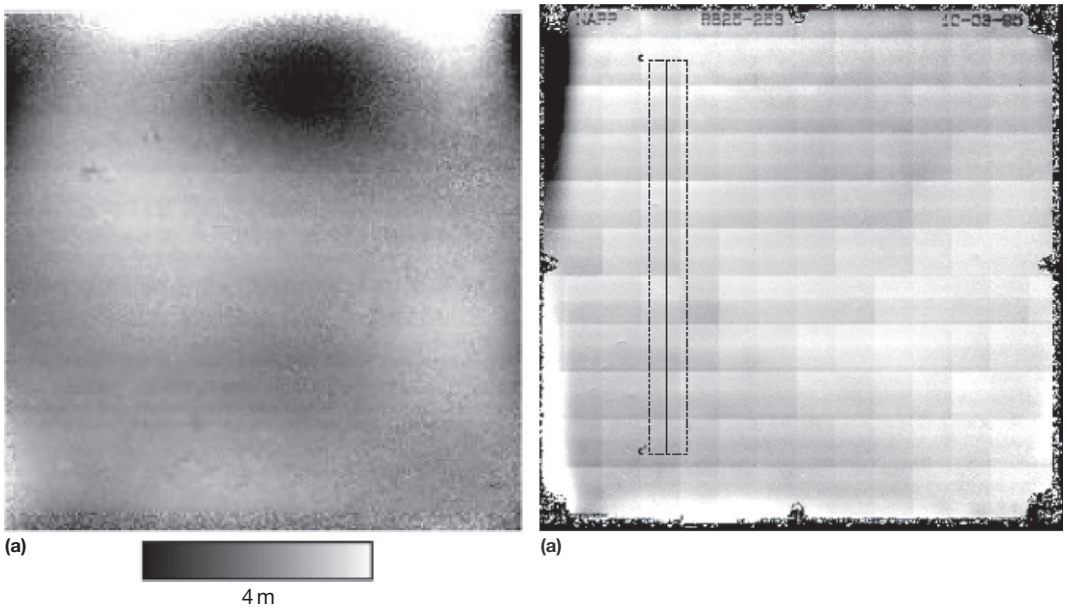


Figure 20 (a) Offset field computed for two air photos from the 1992 Landers earthquake area (US National Aerial Photography Program) (Michel and Avouac, 2006). These images were acquired 3 and 10 years after the earthquake, respectively, and should therefore not show any significant ground deformation. Offsets were measured from the phase correlation technique with a matching window of 32×32 pixels. These images, with a nominal film resolution of 10 mm corresponding to a ground resolution of 0.4 m, were scanned with a microdensitometer designed for astronomy with a theoretical positional accuracy of 0.6 mm and a root-mean-square error (RMSE) of 0.2 mm. The offsets field illustrates typical noise on measurement of ground deformation as measured from aerial photos. The low-frequency pattern results from unrecoverable thermomechanical deformation of films and prevents analysis of deformation of the ground at wavelength larger than about 1 km. Scan artifacts, temporal decorrelations, and shadow effects account for the noise at shorter wavelength. This noise has an rms of about 5 cm. (b) Along-line offsets measured from matching two scans of the same air photo from 1995 with the phase correlation method (Ayoub et al., 2008). One image was scanned with the microdensitomer and is used as the reference image. The other is the digital image downloaded from the US National Aerial Photography Program in 2006 (with a nominal 21 mm resolution). Correlation used a 64×64 window with a 32-pixel step. Scan artifacts are obvious from the typical gridded pattern and reach up to 5 mm (equivalent to 20 cm on the ground). Other long-wavelength deformations are due to film distortion and misregistration.

followed the earthquake (Yu et al., 2001), we consider that the SPOT offsets essentially represent coseismic displacements.

Images acquired with similar near-vertical incidence angles (2.9° and 2.6°) were chosen in order to minimize orthorectification errors. It was assumed that distortions due to topographic errors could be neglected. With this assumption, the horizontal displacement field can be measured from correlating the orthorectified images taken before and after the earthquake. The DEM was produced from digitization of topographic maps with a data spacing of 30 m and an uncertainty estimated to about 20 m RMS. Orthorectification was performed without taking into account the information on the changing attitude of the satellite. Offsets between the two orthorectified images were next computed from the phase shift of the Fourier transform as described in the previous section using a multiscale procedure. They were first computed for a 16×16-pixel correlation window. If the uncertainty on the offset at a particular pixel was found to be larger than 0.3 pixel, the size of the correlation window was locally increased first to 32×32 pixels and then to 64×64 pixels if necessary. Only 5% of the measurements were filtered out by this procedure, yielding independent measurements every 160 m, in general (**Figure 21(a)**). The offsets show the surface rupture but are also clearly affected by residual artifacts with typical >20 km wavelengths (**Figure 21(b)**). These artifacts result from the changing attitude (roll, pitch, and yaw) of the satellite during image acquisition. These effects can now be modeled and optimized following the procedure described in the preceding text. In this early study, they were actually modeled based on the GPS measurements. A total of 59 GPS geodetic measurements were available in the area covered by the SPOT images (see circles in **Figure 21(c)**). The offset field was first filtered using a median filter to remove very short-wavelength (<1 km) noise related to local temporal decorrelation. Differences between offsets and GPS were next calculated for both E–W and N–S components revealing the artifacts introduced by the changing attitude of the satellite. The residuals (**Figure 21(c)**) were then interpolated using bicubic spline functions. These residuals show that jitter effects can bias offset measurements by up to 15 m if they are not properly modeled or compensated for. **Figure 21(d)** shows the amplitude of the offset field after compensation for these biases. This procedure yields a displacement field that benefits from the dense spatial coverage provided by SPOT offsets and the accuracy of the sparse GPS measurements. Jitter effects have been relatively well filtered out. The fault trace is clearly revealed, and the two components of the horizontal slip vector along the fault trace can be measured with an accuracy of a few tenths of centimeters.

This example shows the artifacts potentially introduced by improper modeling of the jitter and how these artifacts can be mitigated when geodetic measurements in the near-field area are available.

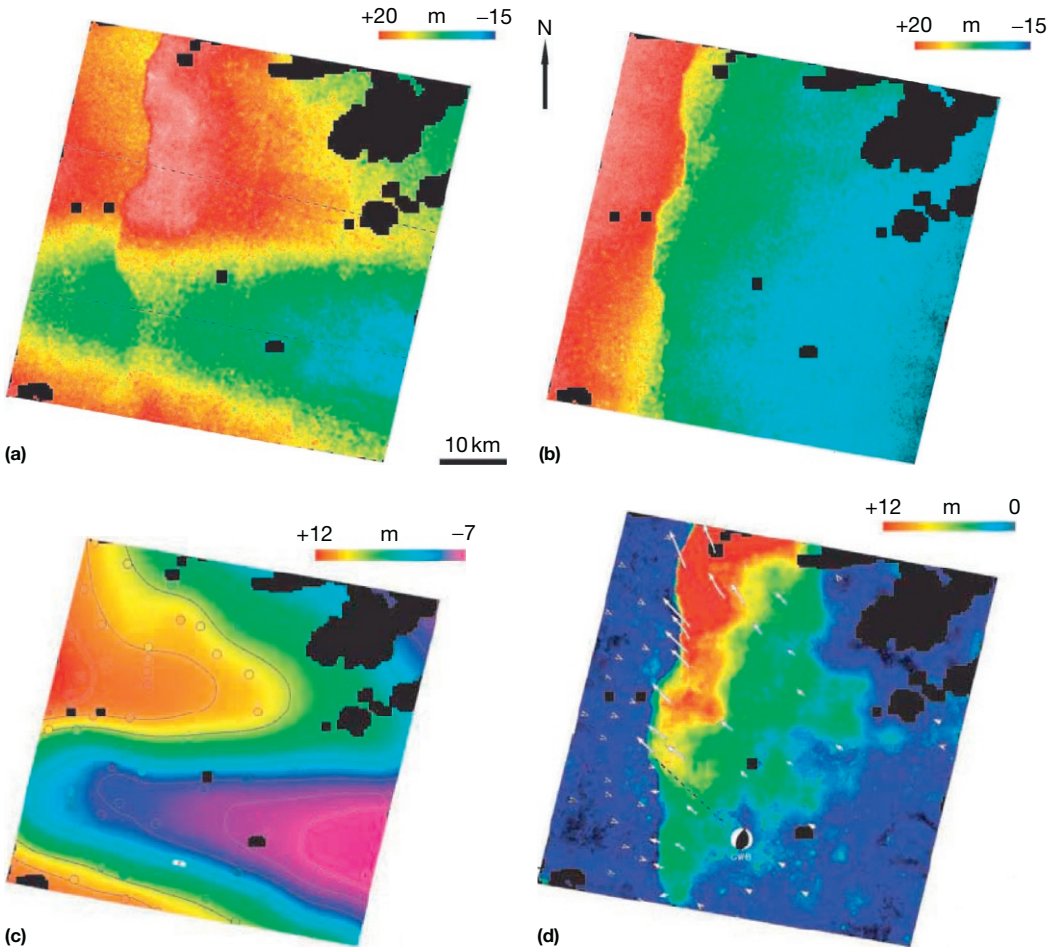


Figure 21 N–S (a) and E–W (b) horizontal offsets determined from the correlation of two SPOT images with 10 m GSD acquired 10 months before and 2 months after the Mw 7.6 Chi-Chi earthquake in 1999. Offsets are determined from the subpixel phase correlation of orthorectified images. Independent measurements every 160 m. The offset field reveals a sharp discontinuity along the fault trace. Long-wavelength artifacts due to the satellite jitter (the changing attitude characterized by the roll, yaw, and pitch) during image acquisition are also obvious especially on the N–S component. These artifacts were modeled based on the residuals (shown in (c)) between those measurements and the GPS measurements collected at 59 sites (circles). Panel (d) shows the amplitude of the offset field after compensation for these residuals filtered at wavelength larger than 20 km. Modified from Dominguez S, et al. (2003) Horizontal coseismic deformation of the 1999 Chi-Chi earthquake measured from SPOT satellite images: Implications for the seismic cycle along the western foothills of central Taiwan. Journal of Geophysical Research-Solid Earth 108(B2), art. no.-2083.

3.13.6.3 Surface Displacement in 2-D due to the 2005 Mw 76 Kashmir Earthquake, Measured from ASTER Images

Another application to a thrust earthquake is illustrated in Figure 22. In that case, two ASTER images (Abrams, 2000) with a ground resolution of 15 m were used to map surface ruptures due to the Mw 7.6 Kashmir earthquake, which struck the northwestern Himalaya near the town of Muzaffarabad on 8 October 2005 (Avouac et al., 2006). This example was particularly challenging due to a number of factors. First, ASTER images have a lower ground resolution of 15 m and less accurate telemetry as compared with the SPOT 2 and SPOT 4 images used in the previous example. The relief in the epicentral area is extremely rugged, enhancing the risk of strong topographic artifacts. Because of the necessity to use images with similar incidence angles so as to limit the topographic errors, two images with quite close view angles were chosen. The first image was acquired on 14 November 2000 and the second on 27 October 2005. The large time span between the two images was enhancing the risk of temporal decorrelation. Finally, there was no geodetic measurement of coseismic displacement in the near-field area, which could have been used to compensate jitter effects.

The images were orthorectified on a common 15 m resolution grid using a 30 m DEM computed from a stereo pair of ASTER images. Offsets were measured from the local cross correlation of the two orthorectified visible near infrared (VNIR) 3 N bands (Figure 22(a)). These measurements show that the surface rupture reached the surface, although field evidence for fault ruptures was scant. In addition to the coseismic signal, a wave pattern, running in the satellite along-track direction, reflects the undersampling of the satellite attitudes that therefore could not be accurately accounted for during orthorectification. This pattern is characteristic of unmodeled pitch variations. A similar pattern on the E–W component reflects roll variations. The Terra onboard attitude

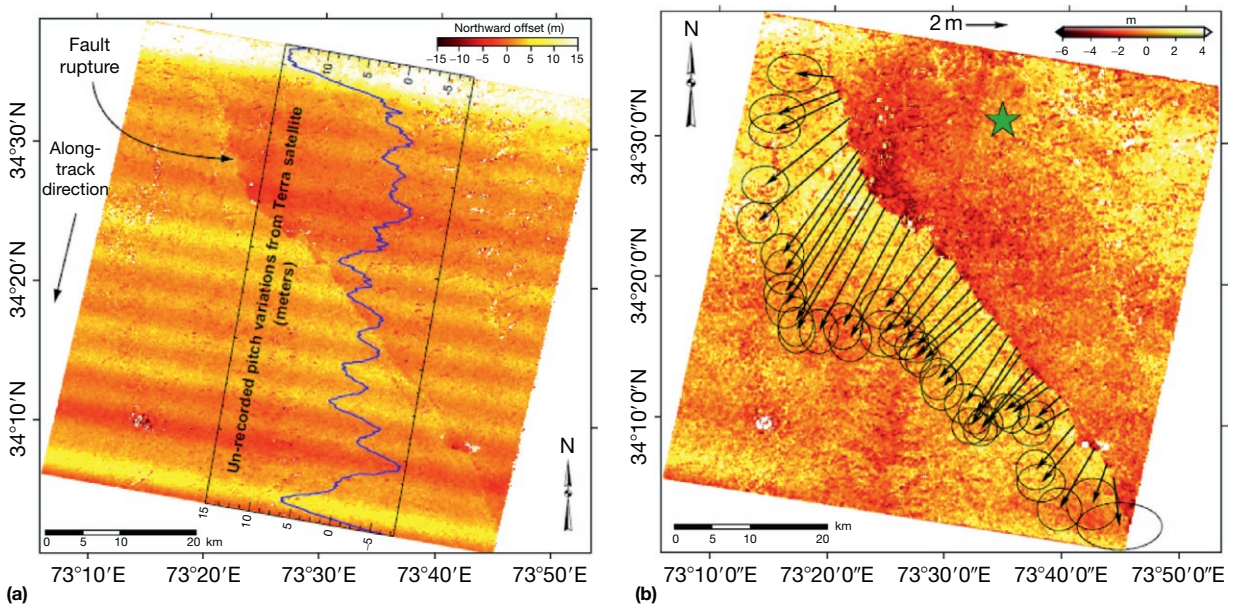


Figure 22 (a) Northward component of offset field measured from correlating ASTER images with 15 m GSD acquired before and after the 2005 Kashmir earthquake. This correlation was obtained with a sliding 32 x 32 pixels correlation window and a 8 pixels step, leading to a ground resolution of 120 m. The points where no measurements were retrieved due to loss of correlation are shown in white. Correlation was lost mostly due to landslides. The fault rupture is visible as a discontinuity in the offset field. The ground displacement signal due to the earthquake is overprint with a wave pattern due to the jitter during the image acquisition. This signal can be estimated by stacking the offset along the lines yielding the blue curve overlaying the offset measurements. (b) Same as (a) after filtering out the jitter effect. Vectors show horizontal slip vectors at about 2 km spacing along the fault trace, measured from the discontinuity of E-W and N-S ground displacement measured at the fault on 18 km long, 6 km wide profiles run perpendicular to the fault. NS and EW offsets at the fault are measured from linear least-squares adjustment on each side of the fault. Ellipses show 2-s uncertainties (95% confidence level) on each measurement.

recorders have a sensitivity of 1 arcsec (corresponding to around 3.4 m on the ground for VNIR images) and a sampling rate of 0.97 Hz corresponding to one measurement every 500 lines or 7.5 km on the ground. This sampling rate is insufficient to record properly the spacecraft jitter observed in Figure 22(a). Similar artifacts in amplitude and frequency were also reported in other studies (Ayoub et al., 2008; Iwasaki and Fujisada, 2003). The relatively short wavelength of the jitter artifacts (5 km, corresponding to vibration around 1 Hz of the Terra platform) would prevent the possibility of compensating them based on geodetic ground measurements as discussed in the previous section. In this particular example, the jitter pattern could be extracted by running profiles in the along-track direction and not intersecting the fault trace. The profiles were next stacked in the across-track direction and subtracted from the original measurements. This yielded the offset field of Figure 22(b) where jitter artifacts are hardly visible any more. This rather crude dejittering procedure thus appears to be quite effective in the context of ASTER images. A more sophisticated optimization procedure could be applied if another image with better controlled attitude was available.

The offset field reveals a clear discontinuity, which can be traced over a distance of about 75 km. Despite the 5-year interval between the two images, the correlation is good, except at locations where major landslides were triggered by the earthquake. The horizontal slip vector on the fault can be measured accurately from profiles run across the fault trace with an accuracy of about 1 m on average at the 95% confidence level. This example shows that images with relatively low ground resolution (15 m in that case) and poor telemetry can nonetheless yield useful measurements.

3.13.6.4 Surface Displacement in 2-D due to the 1999 Mw 7.1 Hector Mine Earthquake Measured from SPOT Images

We now comment on the measurements of coseismic surface displacement due to the Mw 7.1 Hector Mine earthquake in 1999. This earthquake is purely strike slip and a relatively dense measurement of surface fault slip was collected in the field (Treiman et al., 2002) and could be compared with the measurements obtained from optical image correlation. Figure 23 shows the offset field measured from correlating SPOT 4 and SPOT 2 images, with an ~10 m ground resolution, bracketing the earthquake (Leprince et al., 2007). In this example where the two images have a near-vertical incidence angle (<2°), it was assumed that distortions due to topographic errors could be neglected. With this assumption, the horizontal displacement field can be measured from correlating orthorectified images taken before and after the earthquake. Two images taken at about the same date in the year were chosen so that the Sun elevation be as close as possible to avoid shadowing artifacts: a SPOT 4 image (image 1) acquired in August 1998 and a SPOT 2 image (image 2) acquired in August 2000. The two images were orthorectified using the 1 arcsec SRTM DEM (Farr et al., 2007), which has a ground resolution of 30 m.

Correlation was performed with 32×32 pixel (320×320 m^2) sliding windows and with a step of eight pixels (80 m) (Figure 23).

Correlation errors are mostly the consequence of surface changes that occurred during the 2 years separating the acquisitions. Other sources of noise include errors of geometric modeling of the orthoimages due to the topographic errors, CCD misalignments, and mismodeling of the satellite attitude during image acquisition. Altogether, these various sources of errors are estimated to contribute to the noise level with closely unbiased Gaussian distribution and a standard deviation of about 0.8 m on the north–south (N–S) component and 0.57 on the E–W component of displacement. The noise due to correlation errors can be filtered out, for example, using the nonlocal means filter (Buades et al., 2008), which has the advantage of not smoothing out discontinuities. This filter has been applied in the filtered version of Figure 23 shown in Figure 6. The discontinuity due to the surface rupture is therefore well preserved with this filtering technique. The surface rupture appears as a discontinuity in the displacement field. Some systematic errors remain, however. Some are clearly due to CCD misalignments (Figure 23), which were not corrected for with sufficient accuracy. The filtered image (Figure 6) also shows a slight residual jitter effect in the form of a wave pattern in the N–S component of the displacement field. This wave pattern, which is hardly visible in the nonfiltered displacement field (Figure 23), is due to mismodeling of pitch variations and results in offset artifacts with amplitude of about 1 m. These artifacts are a limitation with regard to the measurement of absolute fault displacement. However, they do not affect the measurement of surface fault slip. Due to improvement of telemetry onboard the SPOT program satellite, this kind of artifact is not seen with more recent imaging systems (SPOT 5 and up).

The horizontal slip vector was measured from 8 km long, 880 m wide swath profiles taken perpendicular to the fault trace and spaced every about 880 m (Figure 24). Each horizontal coseismic displacement measured on the fault is up to 6 m in the N–S direction and up to 3.5 m in the E–W direction. The horizontal coseismic fault slip at the surface is therefore accurately and densely (every 80 m) recovered from the proposed technique. With the nominal images resolution of 10 m, all the measurements are in the subpixel range, within ± 3 m. In the N–S correlation image, a secondary rupture branches to the north where the main rupture bends. The coseismic displacement measured on this secondary branch is up to 1 m. These examples show that surface ruptures with surface slip of $< 1/10$ of the pixel size can be detected and measured from this technique.

The location of the fault trace and the surface fault slip recovered from the SPOT image compare well with the surface ruptures and fault slip measured in the field (Figure 24). In fact, fault slip measured from the SPOT images is close to the maximum slip measured in the field and varies smoothly along strike. A similar conclusion has been reached wherever it has been possible to compare field measurement and measurement made to geodetic imaging (Konca et al., 2010; Michel and Avouac, 2002; 2006; Taylor et al., 2008). This observation suggests that the variation of slip generally revealed from field

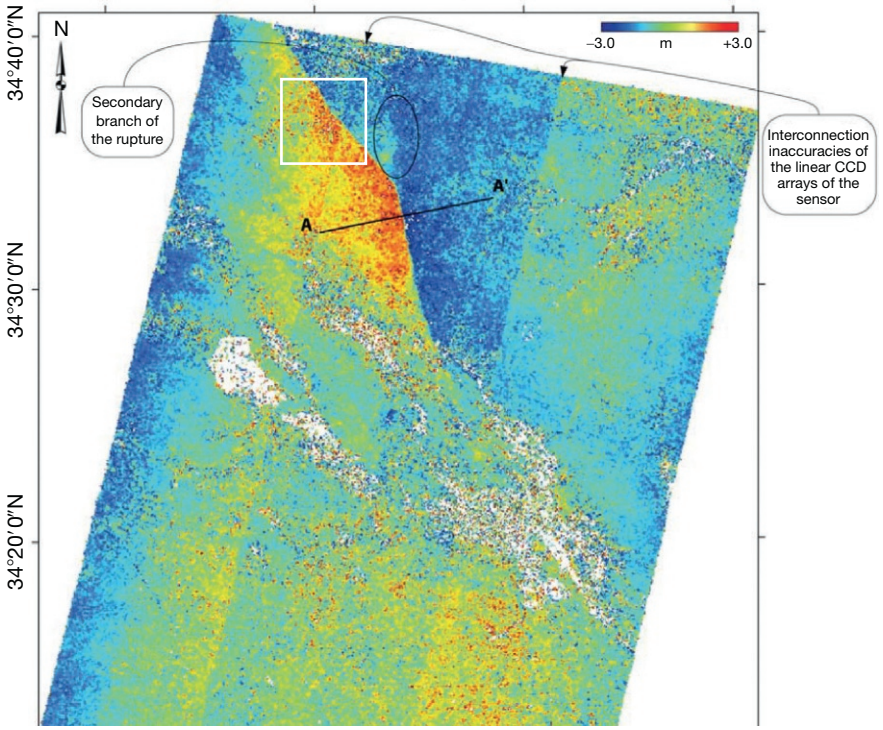


Figure 23 North component (positive to the north) of the coseismic displacement field due to the 1999 Mw 7.1 Hector Mine earthquake in California measured from correlated SPOT 2 and SPOT 4 monochromatic images with 10 m GSD acquired on 12 August 1998 and 10 August 2000 (Leprince et al., 2007). Both images were orthorectified and co-registered on a 10 m-resolution grid, and the offsets were measured from subpixel correlation within a 32×32-pixel sliding window. The displacement field is shown at a ground resolution of 80 m. See Figure 6 for filtered version of these measurements. White box shows location of Figure 25

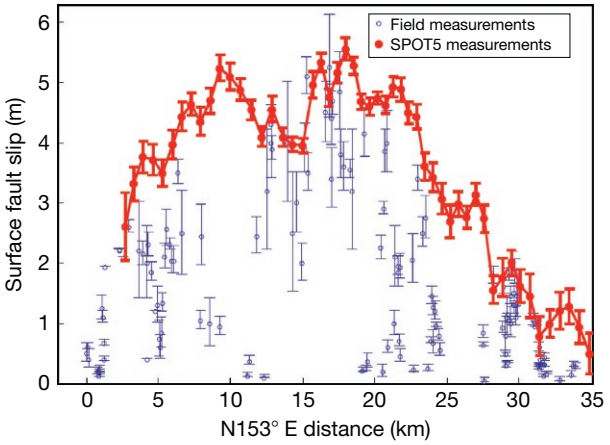


Figure 24 Right-lateral slip along the Hector mine earthquake surface rupture. Slip is determined by projecting the horizontal slip vectors along the fault strike. Horizontal slip vectors are measured from linear least-square adjustment, on each side of the fault and on each N–S and E–W image of stacked profiles running perpendicularly to the rupture (as shown in **Figure 6**). Profiles are stacked over a width of 880 m and a length of 8 km. Error bars show 2-σ uncertainties (95% confidence level). Field measurements (Treiman et al., 2002) are also reported for comparison. Modified from Leprince S, et al. (2007) Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. *IEEE Transactions on Geoscience and Remote Sensing* 45(6): 1529–1558.

investigation of surface rupture at the kilometric scale or less is probably due to the fact that the total slip across the fault is in fact partitioned between slip on faults identifiable in the field and more distributed anelastic strain. This observation suggests that a significant portion of coseismic slip may in some cases escape detection by conventional field measurements alone. A corollary is that the along-strike variability in fault slip observed in the field may actually reflect lateral variations in the mechanical response of near-surface layers to the dynamics of earthquake rupture.

3.13.6.5 Surface Displacement in 2-D due to the 1999 Mw 7.1 Hector Mine Earthquake Measured from Air Photos

The spatial resolution and accuracy of ground displacement resolution scale in principle with the resolution of the optical images. There is nowadays a good archive of satellite imagery with decametric resolution like the SPOT and ASTER images exploited in the examples presented in previous subsections. In principle, these data are appropriate to study Mw > 7 earthquakes, which generally produce surface slips in excess of 1–2 m. The measurement of more subtle deformation would require images with higher ground resolution as now available from a number of satellite programs (e.g., IKONOS, WorldView, and Pleiades). In most countries, air photos with a submetric resolution collected by national surveying agencies are available providing the possibility to revisit past events. The exploitation of such data can be challenging due to scanning artifacts and possible distortions of the prints or films but some success was reached (Ayoub et al., 2009; Michel and Avouac, 2006).

Figure 25 shows, for example, the N–S component of the displacement determined from correlated air photos from the USGS–NAPP (US Geological Survey–National Aerial Photography Program), taken in 1989 and 2002 and covering a portion of the rupture produced by of the Hector Mine earthquake (see Figure 23 for location of footprint). The footprint of these images is about 10×10 km^2. The original film nominal resolution of about 10 μm corresponds to a ground resolution of about 0.4 m.

The two images were orthorectified using the 1 arcsec SRTM RTM DEM. The IO models of the 1989 and 2002 images were established based on the camera calibration reports provided by the USGS. The 2002 image was co-registered first to the topography as the SRTM mission was carried out in 2000 (both postearthquake). A shaded image of the DEM was generated with illumination parameters estimated from the 1989 image shadow pattern. A handful of GCPs were selected between the image and the shaded DEM. Horizontal and vertical coordinates were both obtained from the georeferenced DEM. The average residual misregistration after GCP optimization was estimated to be 2.4 m, while the standard deviation residual was estimated to be 18 m. This latter uncertainty is slightly higher than the 15 m relative horizontal accuracy of the SRTM DEM. More GCPs would have helped improve the co-registration, but the limited radiometric texture due to the desertic setting did not allow it. The 1989 image was orthorectified on a 1 m-resolution grid. Three GCPs, indicated by the black crosses in Figure 25, were optimized to co-register the two images. First, the ground displacement at these locations was ignored. After GCP optimization, the average residual misregistration was evaluated to 1 mm and the standard deviation residual to 35 cm. Ground displacements were next determined from correlating the two co-registered images using a 64×64-pixel sliding window with a 16-pixel step. The discontinuity of ground displacement along the fault trace is nicely revealed. We observed that long-wavelength distortions are introduced to satisfy the no displacement constraints at the location of the GCPs used to co-register the two images. This is obviously an artifact of the procedure. The optimization of the EO model is biased so that the displacements at the location of the GCPs are minimum.

Such artifacts would happen if the footprint of the image is smaller than the deforming zone, as is the case here, and if no a priori information on GCPs displacement is available. In this particular case, the displacement at the location of the GCP can be estimated based on the results obtained from the SPOT images correlation. Panel (b) in Figure 25 shows the results obtained in that case. The long-wavelength distortions have been removed. Only the long-wavelength distortions caused by the SPOT correlation error may remain, along with those introduced by film distortions. These are relatively small and would not impair a precise measurement of fault slip.

This example illustrates one difficulty that generally arises with high-resolution imagery. Due to their limited footprint, it might not be possible to define GCPs outside the deformation zone. If displacements at the location of the GCPs used for the co-registration are unknown, ground deformation at wavelengths larger than the typical wavelength defined by the geographic distribution of the GCPs is filtered out. Deformation at smaller wavelengths is preserved so that this is generally not a

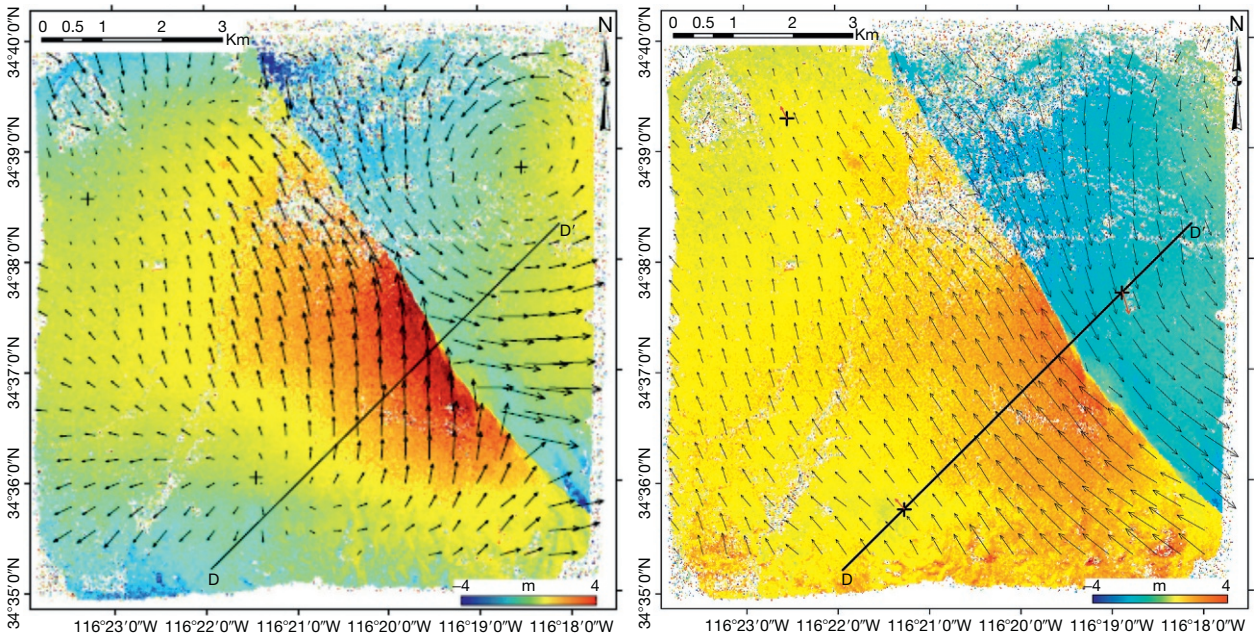


Figure 25 Displacement due to the 1999 Hector Mine earthquake measured from correlation of air photos from 1989 and 2002 with 1 m GSD. The color shade shows amplitude of N–S displacement and vectors show the horizontal displacement vectors. Images were orthorectified on a 1 m grid and correlated using a 64×64-pixel window with a 16-pixel step. Three GCPs, indicated by the black crosses, were optimized to co-register the master and slave images. (a) Long-wavelength distortions are introduced to satisfy the constraint of no displacement at the location of the GCPs during the co-registration. (b) Long-wavelength distortions are removed if the displacements measured from the SPOT image correlation at the location of the GCPs are accounted for during co-registration.

limitation to measure coseismic slip. The general procedure would require a bundle optimization of all the overlapping images. The referencing to a known geodetic reference frame requires that some of the images would extend to non-deforming zones or independent measurements of the displacement at a subset of GCPs.

3.13.6.6 Surface Displacement in 3-D due to the 2010 Mw 7.2 El Mayor–Cucapah Earthquake from LiDAR and Optical Images Stereomatching

The Mw 7.2 El Mayor–Cucapah earthquake that struck southern California on 4 April 2010 is the first earthquake for which pre- and postearthquake LiDAR data were available (Oskin et al., 2012). Pre- and postearthquake images with an ~50 cm GSD and varied viewing angles are also available from WorldView. This exceptional dataset makes it possible to determine the near-field ground displacement in 3-D.

We first show and comment on the 2-D measurements of surface deformation that were produced as early as 1 week after the quake (Wei et al., 2011). These measurements were produced from correlating a pair of panchromatic SPOT 5 images with 2.5 m ground resolution acquired on 26 May 2009 and 8 April 2010, which cover the northern half of the rupture length. Images with near-nadir-looking angles and incidence angle difference of 6° were chosen to minimize topography parallax. These images were orthorectified using the 30 m NED DEM and processed for 2-D horizontal displacement (Figure 26). Matching was performed using the phase correlation method with a 64×64-pixel sliding window at every 16 pixels. It results in a displacement field sampled at 40 m ground resolution. Because of the rugged topography of the Sierra de Los Cucapah and the high resolution of the images compared with the DEM resolution, strong topographic distortions were to be expected. Slight topography residuals indeed show in the E–W component. The N–S component of the displacement field is nearly perpendicular to the epipolar direction of the image pair and is devoid of topography residuals. Both components clearly show the fault trace running through the Sierra de Los Cucapah. It should be pointed out that, due to its unexpected location within the core of the Sierra, these surface ruptures were initially missed during the helicopter reconnaissance carried on in the early days after the quake (Ken Hudnut, personal communication). These measurements reveal a continuous fault trace of about 120 km with an average surface slip of about 2 m. Ruptures with more than 50 cm of surface slip can be detected and mapped with this dataset.

The availability of LiDAR analysis of pre- and postearthquake topographic data provides an opportunity to deliver the full 3-D displacement field of the ground's surface. However, as discussed in Section 3.13.2.1 and illustrated in Figure 2, direct differencing of a pre- and postearthquake digital topography model (DEM) generally leads to biased estimation of the vertical component of the deformation especially if the earthquake also produced significant horizontal motion. To overcome this limitation, we use the COSI-Corr subpixel correlation algorithm to estimate the relative horizontal offset between the pre- and post-2010 El Mayor–Cucapah earthquake high-resolution LiDAR acquisitions. This analysis

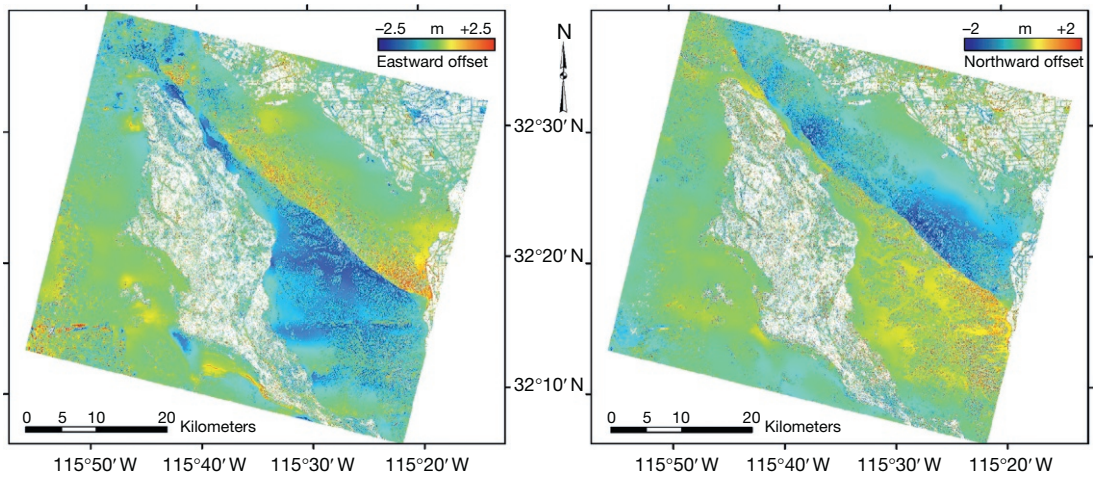


Figure 26 N–S and E–W surface displacements (positive northward) induced by the 2010 Mw 7.2 El Mayor–Cucapah earthquake (Wei et al., 2011). These measurements were produced from a pair of panchromatic SPOT 5 images with 2.5 m GSD acquired on 26 May 2009 and 8 April 2010. These images cover the northern half of the rupture length. These images were orthorectified using the 30 m NED DEM. Image cross correlation was performed using a 64 × 64-pixel sliding window and step of 16 pixels, resulting in a displacement field sampled at every 40 m. Slight topography residuals are visible in the E–W component but the N–S component, which is nearly perpendicular to the epipolar direction of the image pair, is devoid of topography residuals.

shows an unprecedented view of the complete vertical slip component of the rupture induced by the Mw 7.2 2010 El Mayor–Cucapah earthquake, sampled at every 5 m, over a length of about 100 km, and with a vertical accuracy of a few centimeters (Figure 27) (Leprince et al., in preparation). These measurements reveal that the LiDAR data suffer from jitter and tiling artifacts, probably associated to the lower-quality pre-EQ dataset. Horizontal displacements are therefore strongly biased with systematic misregistration errors exceeding several meters. The tectonic signal is clear, however. By contrast, the measurements of vertical displacements are accurate to within 10 cm and reveal nicely details of the fault rupture at the 100 m scale.

We show in Figure 28 the surface displacements retrieved using the 2-D method applied to two WorldView images bracketing the El Mayor–Cucapah earthquake. Both were orthorectified using the 30 m NED DEM. The results show strong topographic artifacts in the N–S component, which is close to the epipolar direction (along the track of the sensor's motion which is mostly in the N–S direction). This example highlights the necessity of stereo acquisitions when images cannot be acquired at nadir.

The results obtained with the 3-D method illustrated in Figure 9 are presented in Figure 29. Four images with 50 cm GSD were used (QuickBird or WorldView). The standard deviation on the measurement of horizontal displacements is about 10 cm (1/5 of the pixel size) and about 25 cm on the vertical. The spatial distribution of horizontal displacements suggests that errors are dominated by CCD and jitter artifacts.

3.13.7 Applications to Geomorphology and Glacier Monitoring

3.13.7.1 Glacier Monitoring

In the current climatic context, the monitoring of continental ice and a better understanding of glaciers' dynamics are crucial as the fate of mountain glaciers is very poorly constrained. This issue can be addressed from cross correlation of optical imagery (Berthier et al., 2005; Heid and Kaab, 2012; Scherler et al., 2008, 2011). The capability of correlation of passive optical images to measure glacier's surface velocities is illustrated in Figure 30. In this application, the technique tracks surface features of the glacier such as crevasses, debris, and ogive bands (Figure 31). These features are assumed to be advectively transported with ice flow. This is a first-order approximation that ignores the effect of ablation and accumulation. The horizontal displacements in the Mer de Glace area (Alps), over 26 days (23 August 2003 to 18 September 2003), were derived from SPOT 5 images with 2.5 m GSD (Leprince et al., 2008a). Very few areas of decorrelation are observed and, when present, are mainly due to changes in length and orientation of the mountain shadows between the two dates. Around the main glaciers, many small disconnected regions (subkilometric size) have measurable motion, then showing a complete picture of the ice flow field. This study reveals details of the ice surface velocity field, which can help calibrate and validate glacier-flow models. The measurements made from the SPOT image compare well with ground-based GPS measurements along a central flow line. The GPS displacements are larger by about 10% probably mostly because they do not cover the exact same time period as the SPOT measurements. The time period covered by the GPS (12 August to 03 September 2003) starts indeed slightly earlier in the summer at the time of the peak heat wave of August 2003 (Berthier et al., 2005). The consistency between the two datasets validates the method despite that some ablation must have occurred in the time period separating the acquisition of the two SPOT images. The measurement was successful most probably because surface features are preserved during ablation as debris remains at the surface and crevasses and ogive bands are bulk properties that extend at depth. As ice is ablated, the surface pattern of crevasses and ogive bands remains approximately stationary.

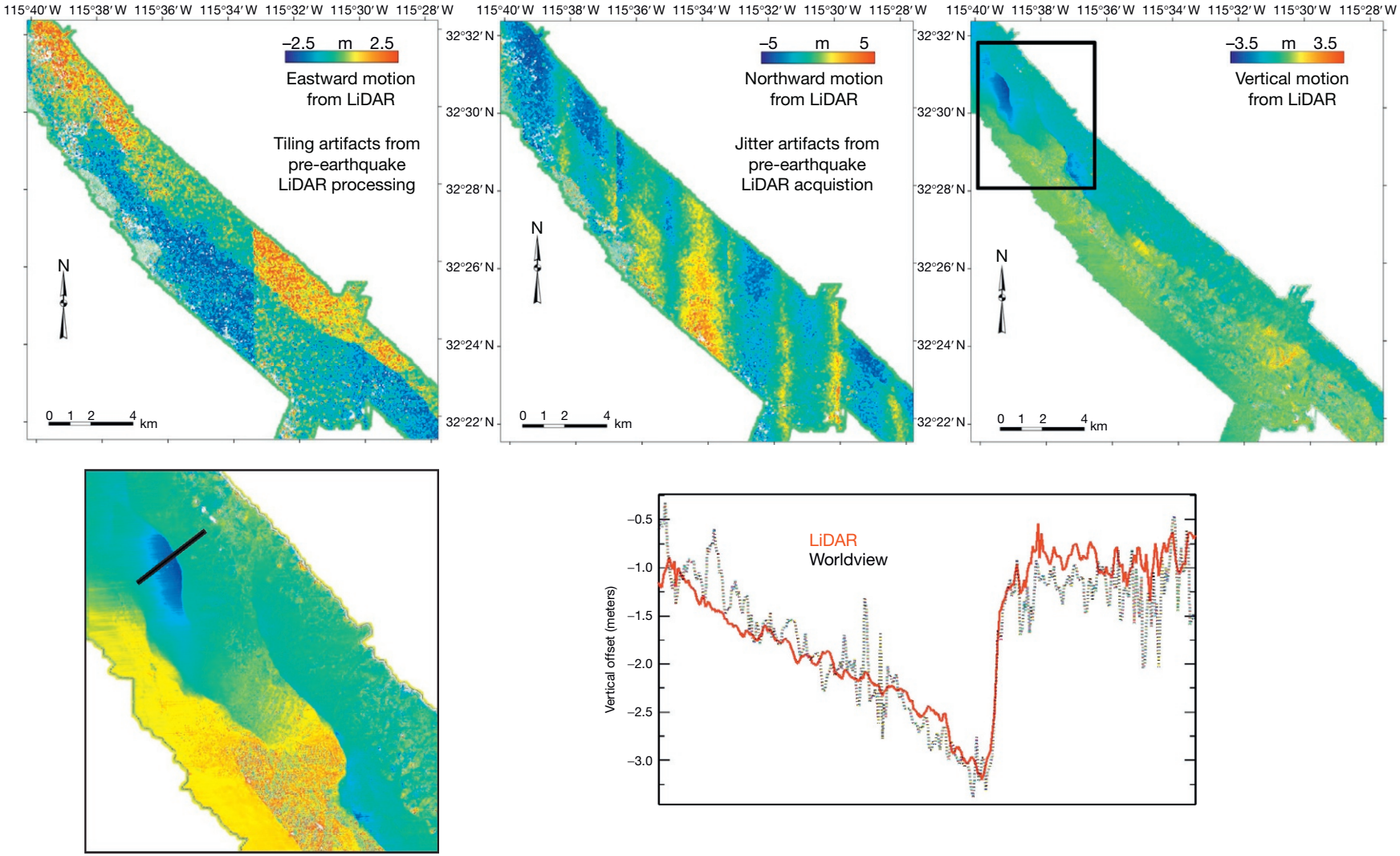


Figure 27 3-D offset recovered from LiDAR and comparison of the vertical offset recovered using WorldView data (**Figure 28**). We see that the LiDAR horizontal components show strong tiling and jitter artifacts due to uncorrected aircraft attitude control and processing inaccuracies. The pre-earthquake LiDAR was gridded at 5 m, and the postearthquake LiDAR, although acquired at a resolution of 50 cm, was resampled on a 5 m grid to match the pre-earthquake LiDAR dataset. Box in upper-right panel shows close-up view of lower left panel where location of profile shown in lower right panel is indicated. Red line shows LIDAR measurement and black dotted line shows World view measurements.

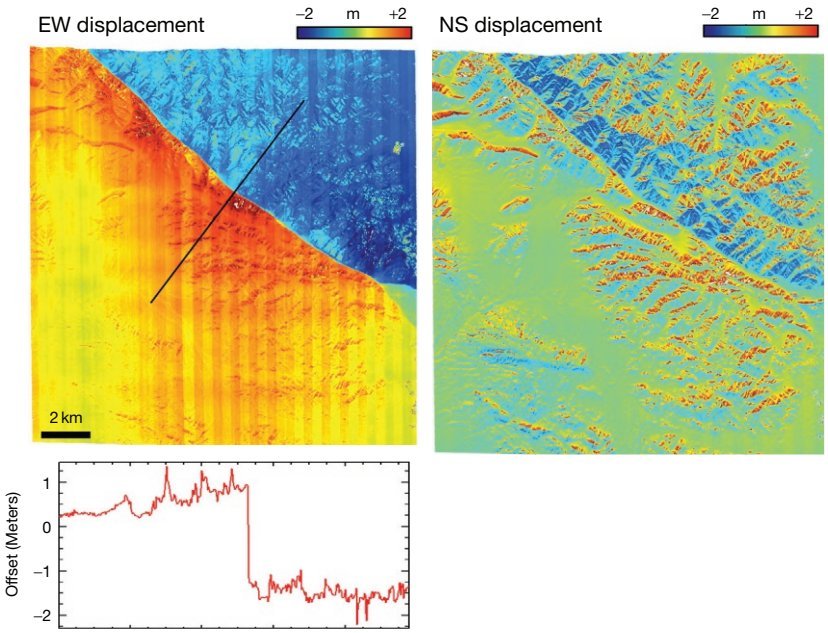


Figure 28 E–W and N–S surface displacements retrieved using only two WorldView images bracketing the El Mayor–Cucapah earthquake. Pre-earthquake images: WorldView 16 September 2008 (along-track angle: −10.8° and across-track angle: 13.5°). Postearthquake images: WorldView 10 April 2011 (along-track angle: −13.8° and across-track angle: −22.5°). We see that large artifacts are introduced in the N–S component, which are due to topographic residuals along the epipolar direction (along-track sensor, which is mostly in the N–S direction). This example highlights the necessity of stereo acquisitions when images cannot be acquired at nadir.

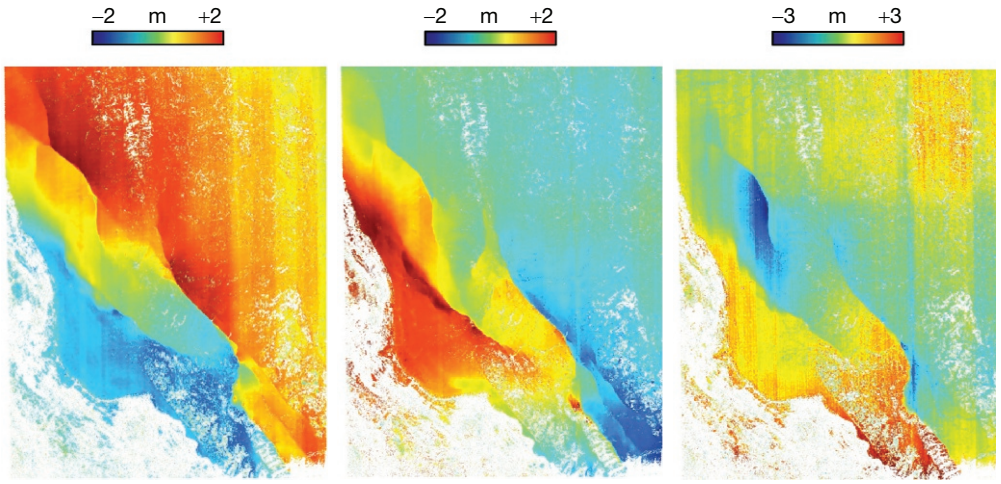


Figure 29 3-D displacement field of the 4 April 2010, El Mayor–Cucapah earthquake retrieved using the method in **Figure 10**. Four images were used: pre-earthquake images, QuickBird 21 September 2006 (along-track angle: −1.23° and across-track angle: −9.8°) and WorldView 16 September 2008 (along-track angle: −10.8° and across-track angle: 13.5°) and postearthquake images, WorldView 10 April 2011 (along-track angle: −13.8° and across-track angle: −22.5°) and WorldView 19 May 2011 (along-track angle: 14.1° and across-track angle 21.6°).

Some bias could be introduced due to the dip angle of the bands. This bias seems negligible here.

This example demonstrates the potential and performance of optical image correlation for the measurement of surface ice flow. In principle, the mass balance of glaciers could be tracked from repeated stereo pairs of optical images or LiDAR survey using the 3-D matching techniques described in the preceding text. Such an approach should be more precise than simple DEM differencing (Berthier et al., 2006) as it should in principle mitigate misregistration bias and provide a better resolution provided that glacier surface features remain coherent in the dataset.

3.13.7.2 Earthflows

Earthflows resulting from slow-moving mass movement are a common phenomenon on clayey hillslopes. Earthflows might be a major geomorphic factor, capable of eroding hillslopes at

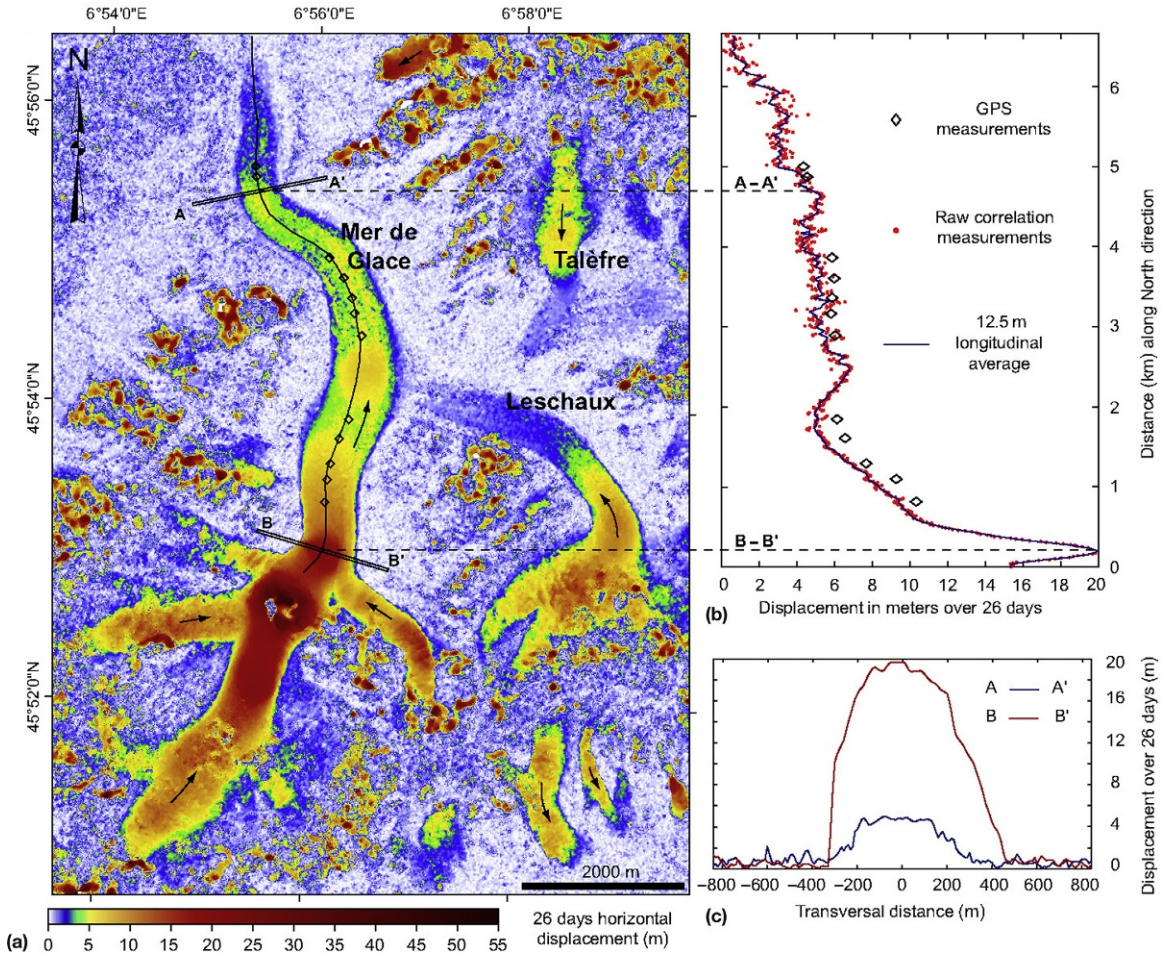


Figure 30 (a) Amplitude of the horizontal displacement over the Mer de Glace area from 23 August 2003 to 18 September 2003 (Leprince et al., 2008a). The displacement field was computed from the subpixel correlation of two panchromatic SPOT 5 images with 2.5 m GSD, using a sliding window of size 32×32 pixels and a step of 16 pixels. Arrows show the flow direction. Displacements as high as 55 m (about 800 m year^{-1}) are recorded over this 26-day period. Box shows location of **Figure 30**. (b) Displacements along a central flow line of the Mer de Glace measured from SPOT 5 images and from GPS campaign measurements. The time period covered by the GPS (12 August 2003 to 03 September 2003) starts slightly earlier in the summer and includes the August 2003 European heat wave, explaining the faster velocities observed over this period (Berthier et al., 2005). (c) Displacements along transverse profiles AA′ and BB′ across the glacier. Displacements were stacked within a 12.5 m wide swath. No topography or baseline artifacts can be noticed.

rates on the order of mm year^{-1} (Mackey and Roering, 2011; Kelsey, 1980). They are also a major natural hazard in mountainous areas. Slow landslides generally have a complex dynamics highly sensitive to climatic factors, which remain poorly understood (Malet et al., 2003). Assessing how they might evolve with time is therefore a challenge. Conventional geodetic measurements (tacheometry, leveling, and GPS geodesy) are commonly used to monitor the temporal evolution of landsliding, but it cannot capture the spatial heterogeneities of mass movement, which may be best assessed using multitemporal optical data or inSAR (e.g., Delacourt et al., 2007; Roering et al., 2009; Travelletti et al., 2012). Displacement rates associated with active earthflows cover a large range from a few mm year^{-1} to a several m h^{-1}. The small scale of these features (<1 km^2) and often large surface strain makes geodetic imaging from optical methods particularly appropriate. As an example, **Figure 32** shows the cumulative horizontal displacement of the La Valette landslide (southern French Alps, Ubaye Valley) over about 11 months, measured from the subpixel correlation of two SPOT 5 images (2.5 m GSD) (Leprince et al., 2008a,b). This displacement field is consistent with inSAR measurements of that same area (Squarzoni et al., 2003) but provides better spatial resolution and resolves displacements not only in the line of sight direction. A network of benchmarks had been settled for repeated geodetic measurements. Even though the targets were correctly placed according to the morphology of the landslide, it turns out that the most active areas revealed by our technique are not covered by the network and otherwise may have remained undetected. Interestingly, the velocity field does not coincide with the geomorphic expression of the landslide and is highly heterogeneous.

Measurement of surface displacements in 3-D is certainly preferable to study earthflows. Based on the mass conservation principle, 3-D measurements can indeed be used to evaluate the geometry of the slip surface at depth (Casson et al., 2005). This information can then be used to determine the volume

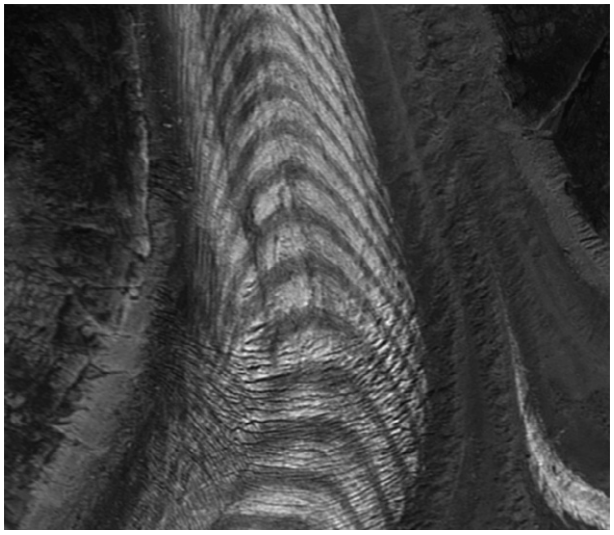


Figure 31 Close-up view of surface features of the Mer de Glace as seen in the SPOT 5 image of 23 August 2003. Note crevasse, ogive bands, and debris cover.

and mass flux involved in the landside or the factors controlling the depth extent and kinematics of the earthflow. Such measurements can be achieved from the stereo matching of optical images described in the preceding text, from 3-D matching of point clouds derived from LiDAR measurements or stereo photogrammetry (Teza et al., 2007), from combining 2-D displacement measured from correlation of optical images, or from shaded DEMs followed by DEM differencing (Casson et al., 2005; Daehne and Corsini, 2012; Teza et al., 2007).

For example, Daehne and Corsini (2012) had used terrestrial LiDAR data with sampling density of about 8–10 point m^{-2} (GSD of 0.3 m) to study a 1.5 km^2 earthflow in the northern Apennines. They first generated DEMs with 1 m posting and measured horizontal displacements from matching slope maps using normalized cross correlation with a 128×128-pixel sliding window. Vertical displacements were determined next from DEM differencing. They were able to measure displacements as large as 60 m. Measurements carried on undeformed areas show that 1-σ uncertainty on the horizontal displacements is on the order of 0.7 m (hence twice the GSD of the original dataset) and also about 0.7 m on the vertical.

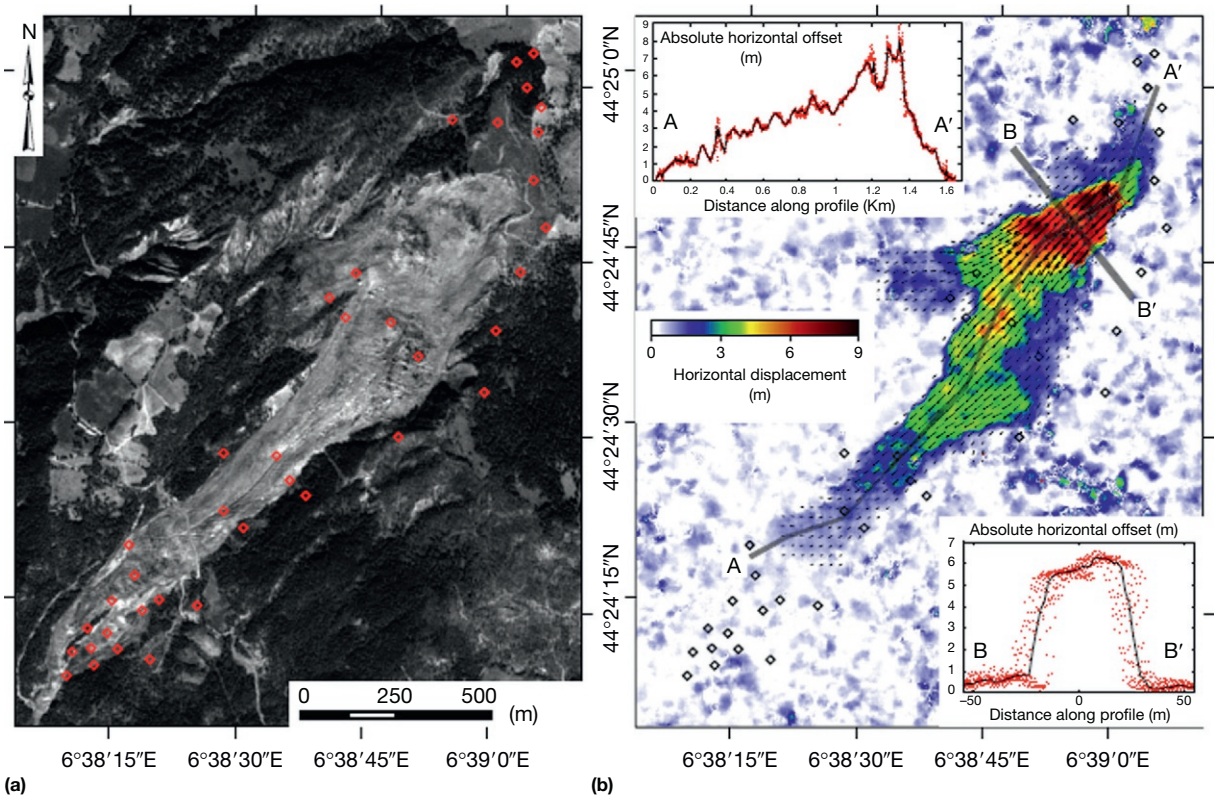


Figure 32 (a) Orthorectified SPOT 5 image of the La Valette landslide area (southern French Alps, Ubaye Valley). Red diamonds show location of geodetic benchmarks for field geodetic measurements. (b) Amplitude of horizontal displacement and displacement vectors determined from the correlation of two SPOT 5 images with two 2.5 m GSD acquired on 19 September 2003 and 22 August 2004. The maximum displacement is 9 m. Longitudinal and transversal profiles, along AA′ and BB′, respectively, show the raw data (red points) and the average over a 15 m wide swath (black line). The displacements revealed from the images would have been unnoticed in the geodetic measurements (black diamonds). Modified from Leprince S, et al. (2008) Monitoring earth surface dynamics with optical imagery. *EOS, Transactions, American Geophysical Union* 89(1).

3.13.7.3 Dune Migration

Geodetic imaging has proven quite efficient to study sand dunes activity (Bridges et al., 2012; Necsoiu et al., 2009; Vermeesch and Drake, 2008). In this application, the signal that is tracked is actually the change of the topography due to erosion of the stoss slope, the dune slope facing incoming winds, and deposition on the lee slope where sand is avalanching from the crest down the 'slip face.' Due to the coupling between winds and topography, the transport of sediment in sand sheet can be seen as the result of migrating bedforms of various scales (Andreotti et al., 2002b). Large-scale dunes are thus often covered with smaller-scale ripples, which migrate at a higher rate (**Figure 33**). In this context, the geodetic imaging does not track ground displacements but variation of reflectance induced by topographic slopes because the albedo is approximately uniform. The images have then to be registered to the underlying bedrock. Precise co-registration requires that the images covered bedrock areas. Note that in principle, not all the images have to contain bedrock features if they can be registered through a bundle adjustment procedure. Hence, we are in the situation represented as in **Figure 2**, but rather than DEM differencing, the technique designed to track ground displacement provides a more appropriate measurement. Image correlation yields the migration rate of the bedforms, which dominate the radiometric texture within the correlation window. This information can then be used directly to compute the sand flux involved in the migration of these bedforms. As an illustration, **Figure 34** shows sand ripple migration rates measured from correlating HiRISE images (with GSD of 25 cm) from the Nili Patera dune field on Mars (Bridges et al., 2012). The insets shows the azimuthal distribution of migration direction, which was found to be consistent with the orientation of the barchans. These measurements show that ripple migration rates increase linearly with elevation over the dune in a manner consistent for the dune morphology to be at steady state. Comparing the sand flux involved in the migration of, respectively, the ripples and the whole dunes allowed determining the relative proportion of low-energy sand grains hopping over short distances, the 'reptons,' and the higher-energy 'saltons' that do not contribute much to ripple migration but are involved in whole dune migration. Finally, by comparing the dune volume with the sand flux or the dune width with migration rates, we were able to estimate the turnover time (the time needed for a complete remobilization of its sand). This is a key quantity to characterize dune activity, which is generally measured on Earth from luminescence or cosmogenic dating (e.g., Bristow et al., 2007; Singhvi and Porat, 2008; Vermeesch et al., 2010).

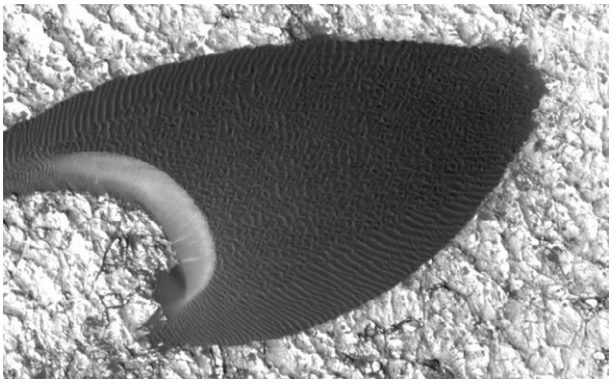


Figure 33 Close up view of an HiRISE image of a barchan dune at Nili Patera. These images have a GSD of 25 cm, sufficient to resolve the ripples, which adorn the dunes.

This example, together with a number of other studies of dune migration using remote sensing (Hugenholtz et al., 2012), demonstrates the potential of these techniques to monitor eolian activity. This is of major interest as the data provided through these techniques would help address fundamental questions in eolian geomorphology (Bourke et al., 2010; Titus et al., 2008). Such data would in particular allow validation and calibration of numerical models of sand transport and dune dynamics (e.g.,Claudin and Andreotti, 2006; Elbelrhiti et al., 2008; Fenton and Richardson, 2001; Kok and Renno, 2009; Richardson et al., 2007); they would complement geologic indicators of dune dynamics to help infer the time evolution of eolian bedforms and past climate change (Beveridge et al., 2006; Hayward et al., 2009) and help assess the impact of current climate change on dune activity (in relation to wind regime, vegetation, and humidity changes) and atmospheric dust. They would also improve our understanding of how various dune forms (e.g., simple and compound barchans and transverse and longitudinal linear dunes) and their morphometric characteristics emerge and evolve due to their interactions and in relation to the wind regime and boundary conditions (Andreotti et al., 2002a,b; Bristow et al., 2007; Ewing and Kocurek, 2010; Reffet et al., 2010).

3.13.8 Conclusion

Thanks to the increasing availability of LiDAR dataset, high-quality optical satellite images, and advances in optical image modeling and matching techniques, it is now possible to detect and measure accurately changes of the Earth's surface due to a variety of processes. These techniques cannot compete with conventional geodesy with regard to the accuracy of pointwise measurements, but they provide a spatial coverage that would be impossible to match with ground-based techniques.

These techniques are sufficiently mature that they can be used to address a variety of topics in Earth sciences and for disaster mitigation. They can be used in particular to provide rapid information on ground deformation and damages in the epicentral area of large earthquakes, to measure the Earth's surface changes due to catastrophic and slow-moving landslides, to monitor ice flow and the mass balance of mountain glaciers, or to monitor eolian sand transport. There are a number of limiting factors: passive optical imaging works only at daytime under clear sky condition to limit occlusions by clouds; misregistration due to jitter, CCD and scan artifacts; image matching failure; view angles must differ by less than about 10° or images must be augmented with stereo acquisitions to compensate parallax artifacts.

Various techniques can be used to match passive optical images and generally allow determining horizontal offsets with an accuracy better than 1/10th the pixel size with typically a

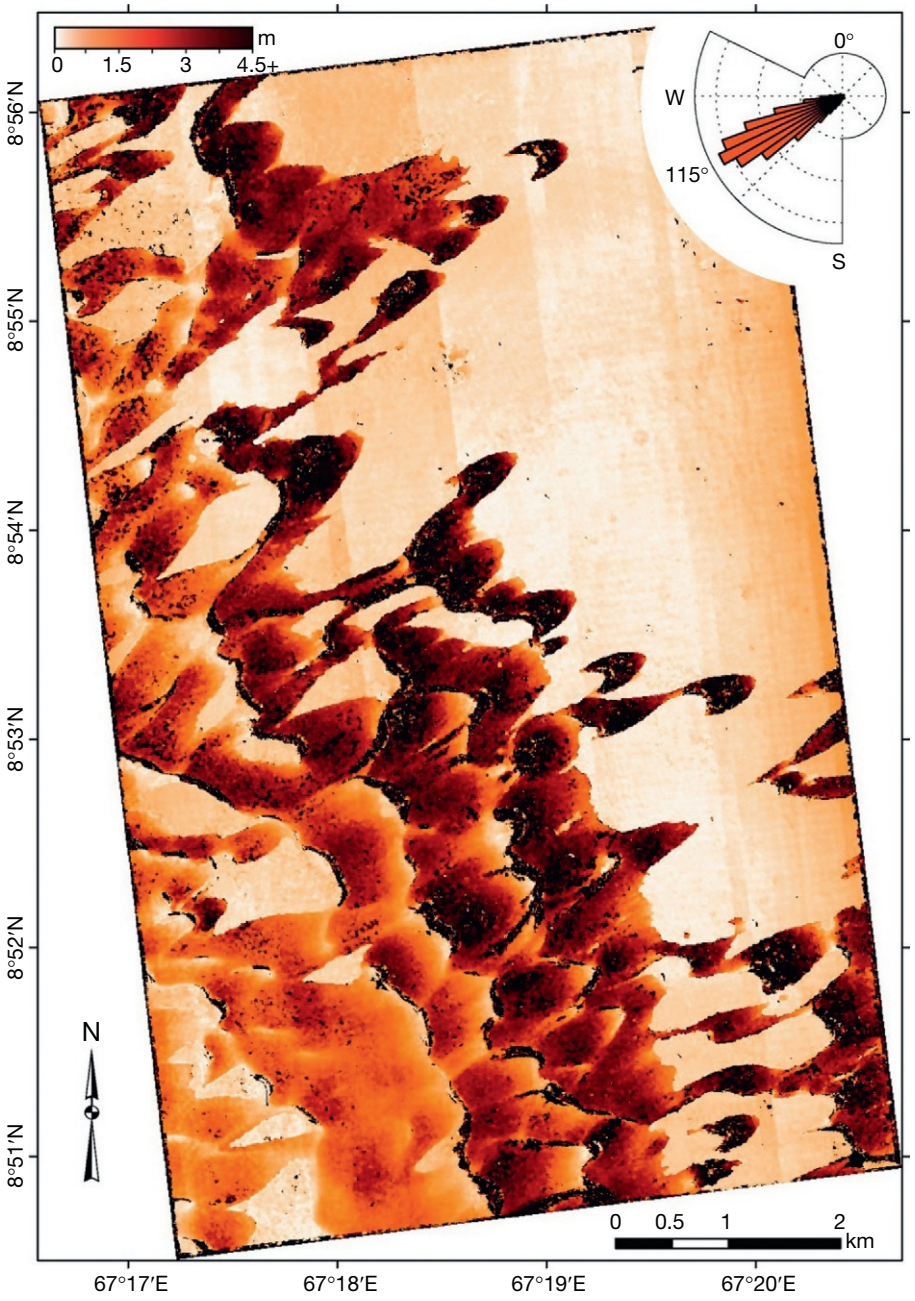


Figure 34 Amplitude of sand ripple migration in the Nili Patera dune field on Mars obtained from the correlation of two HiRISE images 3 months apart (Bridges et al., 2012). The wind rise in inset represents the azimuthal distribution of ripple displacement vectors. Box shows location of HiRISE close-up view shown in **Figure 32**

32×32-pixel window and images with ground sampling of 1m and lower, and view angles differing by less than about 10°. These techniques perform less well on images with submetric resolution and view angles differing by more than 10° as disparity gradients tend to be larger and more variable locally especially with rugged topography or in urban areas. The assumptions that the transformation field can be approximated by a homogeneous rotation globally combined with a locally homogeneous translation (at the scale of the matching window generally used for regularization) does not hold anymore. Geodetic imaging would therefore benefit from more sophisticated matching techniques. It would improve the performance achievable with high-resolution (<1 m) images or with lower-resolution images but large difference of view angles (>10).

Thanks to improved attitude control of most modern satellite imaging systems and the optimization techniques described in this chapter, jitter artifacts have been reduced considerably to be typically subpixel. These artifacts are often the major source of errors over areas with good correlation.

CCD artifacts can in principle be measured and calibrated with accuracy on the order of the percent of the pixel size, but this procedure is heavy and requires appropriate calibration datasets.

Repeating LiDAR measurements offer the possibility to measure vertical displacement with <10 cm accuracy. Horizontal displacements might also be measured from 2-D or 3-D matching of the point cloud but errors can be as large, as the GSD due to misregistrations (mostly jitter during image acquisition). It is hard to correct for these artifacts in a post-processing step. Controlling the orientation model of LiDAR acquisition is therefore required to improve the accuracy of geodetic imaging with such data. Measurements in 3-D are best achieved with passive optical imaging at the moment using a full 3-D co-registration and correlation scheme described in this chapter. Future efforts should focus on improving the capability of these techniques for 3-D matching, the development of more robust matching algorithms that would allow matching accurately high-resolution images.

The temporal resolution of geodetic imaging from optical techniques is currently limited. Optical imaging from satellite is generally Sun-synchronous and on near-polar orbits leading to a typical revisit time at nadir on the order of several months. Even with steering capability, the repeatability is limited as high incidence angles viewed are impaired by occlusions. Repeatability is improved when images from different imaging systems can be combined. This approach requires matching to be performed on ground-projected images following the procedures described in this chapter. In the future, higher temporal resolution will be available as the number of Earth-observing optical imaging system increases. In principle, such systems might be used opportunistically to measure very transient deformation. For example, it has already been proved that oceanic wave propagation can be tracked from current systems, taking advantage of the fact that the CCD arrays of multispectral imaging systems are generally offset in the focal plane of the telescope and image the same spot on the ground at different times (de Michele et al., 2012). It is probable that video imaging system will become operational in the near future opening the possibility for more systematic monitoring of transient phenomena such as seismic waves or catastrophic landslides (Michel et al., 2013).

References

Abrams M (2000) The advanced spaceborne thermal emission and reflection radiometer (ASTER): Data products for the high spatial resolution imager on NASA's Terra platform. *International Journal of Remote Sensing* 21(5): 847–859.

Altamimi Z, et al. (2002) ITRF2000: A new release of the International Terrestrial Reference frame for earth science applications. *Journal of Geophysical Research-Solid Earth* 107(B10).

Andreotti B, et al. (2002a) Selection of dune shapes and velocities—Part 2: A two-dimensional modelling. *European Physical Journal B* 28(3): 341–352.

Andreotti B, et al. (2002b) Selection of dune shapes and velocities—Part 1: Dynamics of sand, wind and barchans. *European Physical Journal B* 28(3): 321–339.

Arnold K (1967) Use of satellites for geodetic studies. *Space Science Reviews* 7(1): 4–68.

Arrowsmith JR and Zielke O (2009) Tectonic geomorphology of the San Andreas Fault zone from high resolution topography: An example from the Cholame segment. *Geomorphology* 113(1–2): 70–81.

Aryal A, et al. (2012) Displacement fields from point cloud data: Application of particle imaging velocimetry to landslide geodesy. *Journal of Geophysical Research-Earth Surface* 117.

Avouac JP, et al. (2006) The 2005, M-w 7.6 Kashmir earthquake: Sub-pixel correlation of ASTER images and seismic waveforms analysis. *Earth and Planetary Science Letters* 249(3–4): 514–528.

Ayoub F, et al. (2008) Influence of camera distortions on satellite image registration and change detection applications. In:*Proc. IGARSS, Boston, MA, USA, July 2008*.

Ayoub F, et al. (2009) Co-registration and correlation of aerial photographs for ground deformation measurements. *ISPRS Journal of Photogrammetry and Remote Sensing* 64(6): 551–560.

Baltsavias EP (1999) Airborne laser scanning: Basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing* 54(2–3): 199–214.

Barnea DI and Silverma HF (1972) Class or algorithms for fast digital image registration. *IEEE Transactions on Computers, C* 21(2): 179–186.

Beckouche S, et al. (2011) Robust Outliers Detection in Image Point Matching.

Bernard S, et al. (2007) Kinematics of fault-related folding derived from a sandbox experiment. *Journal of Geophysical Research-Solid Earth* 112(B3).

Berthier E, et al. (2005) Surface motion of mountain glaciers derived from satellite optical imagery. *Remote Sensing of Environment* 95(1): 14–28.

Berthier E, et al. (2006) Biases of SRTM in high-mountain areas: Implications for the monitoring of glacier volume changes. *Geophysical Research Letters* 33(8).

Besl PJ and McKay ND (1992) A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2): 239–256.

Beveridge C, et al. (2006) Development of spatially diverse and complex dune-field patterns: Gran Desierto Dune Field, Sonora, Mexico. *Sedimentology* 53(6): 1391–1409.

Bevis M, Hudnut K, Sanchez R, et al. (2005) The B4 project: Scanning the San Andreas and San Jacinto fault zones. *EOS Transactions AGU* 86: 52, abstract H34B-01.

Binet R and Bollinger L (2005) Horizontal coseismic deformation of the 2003 Bam (Iran) earthquake measured from SPOT-5 THR satellite imagery. *Geophysical Research Letters* 32: L02307. http://dx.doi.org/10.1029/2004GL021897.

Bishop MP, et al. (2012) Geospatial technologies and digital geomorphological mapping: Concepts, issues and research. *Geomorphology* 137(1): 5–26.

Borsa A and Minster JB (2012) Rapid determination of near-fault earthquake deformation using differential LiDAR. *Bulletin of the Seismological Society of America* 102(4): 1335–1347.

Bourke MC, et al. (2010) Extraterrestrial dunes: An introduction to the special issue on planetary dune systems. *Geomorphology* 121(1–2): 1–14.

Brecher HH (1986) Surface velocity determination on large polar glaciers by aerial photogrammetry. *Annals of Glaciology* 8: 22–26.

Bridges NT, et al. (2012) Earth-like sand fluxes on Mars. *Nature* 485(7398): 339–342.

Bristow CS, et al. (2007) Age and dynamics of linear dunes in the Namib Desert. *Geology* 35(6): 555–558.

Broxton MJ, et al. (2009) 3D Lunar Terrain Reconstruction from Apollo Images. In: Bebis G, et al. (eds.) *Advances in Visual Computing*, pp. 710–719. Berlin/Heidelberg: Springer, Proceedings, Part 1.

Buades A, et al. (2008) Nonlocal image and movie denoising. *International Journal of Computer Vision* 76(2): 123–139.

Burbank D and Anderson R (2001) *Tectonic Geomorphology*, p. 274. Oxford: Blackwell Science.

Carter WE, et al. (2007) Geodetic laser scanning. *Physics Today* 60(12): 41–47.

Casson B, et al. (2005) Contribution of multi-temporal remote sensing images to characterize landslide slip surface—Application to the La Clapiere landslide France). *Natural Hazards and Earth System Sciences* 5(3): 425–437.

Chen Y and Medioni G (1992) Object modeling by registration of multiple range images. *Image and Vision Computing* 10(3): 145–155.

Chen Y-G, et al. (2001) Surface rupture of 1999 Chi-Chi earthquake yields insights on active tectonics of Central Taiwan. *Bulletin of the Seismological Society of America* 91(5): 977–985.

Chini M, et al. (2011) Co-seismic surface effects from very high resolution panchromatic images: The case of the 2005 Kashmir (Pakistan) earthquake. *Natural Hazards and Earth System Sciences* 11(3): 931–943.

Claudin P and Andreotti B (2006) A scaling law for aeolian dunes on Mars, Venus, Earth, and for subaqueous ripples. *Earth and Planetary Science Letters* 252(1–2): 30–44.

Copley A, Avouac JP, Hollingsworth J, and Leprince S (2011) The 2001 M(w) 7.6 Bhuj earthquake, low fault friction, and the crustal support of plate driving forces in India. *Journal of Geophysical Research-Solid Earth* 116. http://dx.doi.org/10.1029/2010jb008137.

Copley A, et al. (2012) Constraints on fault and lithosphere rheology from the coseismic slip and postseismic afterslip of the 2006 M(w)7.0 Mozambique earthquake. *Journal of Geophysical Research-Solid Earth* 117.

Corsini A, et al. (2009) Estimating mass-wasting processes in active earth slides—Earth flows with time-series of high-resolution DEMs from photogrammetry and airborne LiDAR. *Natural Hazards and Earth System Sciences* 9(2): 433–439.
Crippen RE (1992) Measurement of subresolution terrain displacements using SPOT panchromatic imagery. *Episodes* 15(1): 56–61.
Daehne A and Corsini A (2012) Kinematics of active earthflows revealed by digital image correlation and DEM subtraction techniques applied to multi-temporal LiDAR data. *Earth Surface Processes and Landforms* 38: 640–654.
de Michele M, et al. (2012) Direct measurement of ocean waves velocity field from a single SPOT-5 dataset. *Remote Sensing of Environment* 119: 266–271.
Debella-Gilo M and Kaab A (2011) Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. *Remote Sensing of Environment* 115(1): 130–142.
Debella-Gilo M and Kaab A (2012) Measurement of surface displacement and deformation of mass movements using least squares matching of repeat high resolution satellite and aerial images. *Remote Sensing* 4(1): 43–67.
Delacourt C, et al. (2007) Remote-sensing techniques for analysing landslide kinematics: A review. *Bulletin de la Societe Geologique de France* 178(2): 89–100.
Dominguez S, et al. (2003) Horizontal coseismic deformation of the 1999 Chi-Chi earthquake measured from SPOT satellite images: Implications for the seismic cycle along the western foothills of central Taiwan. *Journal of Geophysical Research-Solid Earth* 108(B2), art. no.-2083.
Dudderar TD and Simpkins PG (1977) Laser speckle photography in a fluid medium. *Nature* 270(5632): 45–47.
Elbelrhiti H, et al. (2008) Barchan dune corridors: Field characterization and investigation of control parameters. *Journal of Geophysical Research-Earth Surface* 113(F2).
Ewing RC and Kocurek G (2010) Aeolian dune-field pattern boundary conditions. *Geomorphology* 114(3): 175–187.
Farr TG, et al. (2007) The shuttle radar topography mission. *Reviews of Geophysics* 45(2).
Fenton LK and Richardson MI (2001) Martian surface winds: Insensitivity to orbital changes and implications for aeolian processes. *Journal of Geophysical Research-Planets* 106(E12): 32885–32902.
Fischler MA and Bolles RC (1981) Random sample consensus—A paradigm for model-fitting with applications to image-analysis and automated cartography. *Communications of the ACM* 24(6): 381–395.
Forstner W and Gulich E (1987) A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: *Proceedings of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data,* Interlaken.
French JR (2003) Airborne LiDAR in support of geomorphological and hydraulic modelling. *Earth Surface Processes and Landforms* 28(3): 321–335.
Grenander U and Miller MI (1998) Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics* 56(4): 617–694.
Hayward RK, et al. (2009) Aeolian dunes as ground truth for atmospheric modeling on Mars. *Journal of Geophysical Research-Planets* 114.
Heid T and Kaab A (2012) Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery. *Remote Sensing of Environment* 118: 339–355.
Hild F and Roux S (2006) Digital image correlation: From displacement measurement to identification of elastic properties—A review. *Strain* 42(2): 69–80.
Hirschmuller H (2005) Accurate and efficient stereo processing by semi-global matching and mutual information. In: Schmid C, et al. (eds.), *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, Proceedings,* pp. 807–814.
Hirschmuller H and Scharstein D (2009) Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(9): 1582–1599.
Horn BKP and Schunck BG (1981) Determining optical flow. *Artificial Intelligence* 17: 185–203.
Hudnut KW, et al. (2002) High-resolution topography along surface rupture of the 16 October 1999 Hector Mine, California, earthquake (M-w 7. 1) from airborne laser swath mapping. *Bulletin of the Seismological Society of America* 92(4): 1570–1576.
Hugenholtz CH, et al. (2012) Remote sensing and spatial analysis of aeolian sand dunes: A review and outlook. *Earth-Science Reviews* 111(3–4): 319–334.
Iwasaki A and Fujisada H (2003) Image correlation tool for ASTER geometric validation. In: Fujisada H, et al. (eds.) *Sensors, Systems and Next-Generation Satellites VI*, pp.111–120, SPIE Press.
Jensen JR (2006) *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River, NJ: Prentice Hall.
Kääb A, et al. (1997) Analysing the creep of mountain permafrost using high precision aerial photogrammetry: 25 years of monitoring Gruben Rock Glacier, Swiss Alps. *Permafrost and Periglacial Processes* 8(4): 409–426.
Karabacak VKV, et al. (2011) Monitoring aseismic surface creep along the North Anatolian Fault (Turkey) using ground-based LIDAR. *Earth and Planetary Science Letters* 304(1–2): 64–70.
Kelsey HM (1980) A sediment budget and an analysis of geomorphic process in the van-Duzen River Basin, North Coastal California, 1941–1975 – Summary. *Geological Society of America Bulletin* 91(4): 190–195.
Kirby E and Whipple KX (2012) Expression of active tectonics in erosional landscapes. *Journal of Structural Geology* 44: 54–75.
Klinger Y, et al. (2006) Evidence for an earthquake barrier model from Mw similar to 7.8 Kokoxili (Tibet) earthquake slip-distribution. *Earth and Planetary Science Letters* 242(3–4): 354–364.
Kok JF and Renno NO (2009) A comprehensive numerical model of steady state saltation (COMSALT). *Journal of Geophysical Research-Atmospheres* 114. http/dx.doi.org/10.1029/2009jd011702.
Konca AO, et al. (2010) Rupture Process of the 1999 M-w 7.1 Duzce earthquake from joint analysis of SPOT, GPS, InSAR, strong-motion, and teleseismic data: A supershear rupture with variable rupture velocity. *Bulletin of the Seismological Society of America* 100(1): 267–288.
Kovalevsky J and Barlier F (1967) Geodesie terrestre et geodesie par satellites. *Space Science Reviews* 7(1): 69–134.
Krabill WB, et al. (1984) Airborne laser topographic mapping results. *Photogrammetric Engineering and Remote Sensing* 50(6): 685–694.
Laussedat A (1854) Mémoire sur l'emploi de la chambre claire dans les reconnaissances topographiques. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* 16.
Laussedat A (1859) Mémoire sur l'emploi de la photographie dans la levée des plans. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* 49: 732–734.
Leprince S, et al. (2007) Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. *IEEE Transactions on Geoscience and Remote Sensing* 45(6): 1529–1558.
Leprince S, et al. (2008a) Monitoring earth surface dynamics with optical imagery. *Eos, Transactions American Geophysical Union* 89(1).
Leprince S, et al. (2008b) In-flight CCD distortion calibration for pushbroom satellites based on subpixel correlation. *IEEE Transactions on Geoscience and Remote Sensing* 46(9): 2675–2683.
Leprince S, et al. (2010) Rigorous adaptative resampling for high resolution image warping. In: *IGARSS: 2010 IEEE International Geoscience and Remote Sensing Symposium,* pp. 1943–1946.
Low KL and Lastra A (2003) Reliable and rapidly-converging ICP algorithm using multiresolution smoothing. In: Kawada S (ed.) *Fourth International Conference on 3-D Digital Imaging and Modeling, Proceedings*, pp 171–178.
Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2): 91–110.
Lucas D and Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: *IJCAI Proceedings of the 7th International Conference on Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers.
Lucchitta BK and Ferguson HM (1986) Antarctica—Measuring glacier velocity from satellite images. *Science* 234(4780): 1105–1108.
Mackey BH and Roering JJ (2011) Sediment yield, spatial characteristics, and the long-term evolution of active earthflows determined from airborne LiDAR and historical aerial photographs, Eel River, California. *Geological Society of America Bulletin* 123(7–8): 1560–1576.
Mackey BH, et al. (2009) Long-term kinematics and sediment flux of an active earthflow, Eel River, California. *Geology* 37(9): 803–806.
Malet J-P, et al. (2003) Soil surface characteristics influence on infiltration in black marls: Application to the Super-Sauze earthflow (Southern Alps, France). *Earth Surface Processes and Landforms* 28(5): 547–564.
Marcello J, et al. (2008) Motion estimation techniques to automatically track oceanographic thermal structures in multisensor image, sequences. *IEEE Transactions on Geoscience and Remote Sensing* 46(9): 2743–2762.
Massonnet D and Feigl KL (1998) Radar interferometry and its application to changes in the Earth's surface. *Reviews of Geophysics* 36(4): 441–500.
McKean J and Roering J (2004) Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology* 57(3–4): 331–351.

Michel R and Avouac JP (2002) Deformation due to the 17 August 1999 Izmit, Turkey, earthquake measured from SPOT images. *Journal of Geophysical Research-Solid Earth* 107(B4), art. no. 2062.

Michel R and Avouac JP (2006) Coseismic surface deformation from air photos: The Kickapoo step over in the 1992 Landers rupture. *Journal of Geophysical Research-Solid Earth* 111(B3).

Michel R, et al. (1999) Measuring near field coseismic displacements from SAR images: Application to the Landers earthquake. *Geophysical Research Letters* 26(19): 3017–3020.

Michel R, et al. (2013) A geostationary optical seismometer, proof of concept. *IEEE Transactions on Geoscience and Remote Sensing* 51(1): 695–703.

Necsoiu M, et al. (2009) Monitoring migration rates of an active subarctic dune field using optical imagery. *Remote Sensing of Environment* 113(11): 2441–2447.

Nissen E, et al. (2012) Three-dimensional surface displacements and rotations from differencing pre- and post-earthquake LiDAR point clouds. *Geophysical Research Letters* 39.

Okada Y (1985) Surface deformation to shear and tensile faults in a half space. *Bulletin of the Seismological Society of America* 75: 1135–1154.

Oppikofer T, et al. (2009) Characterization and monitoring of the Aknes rockslide using terrestrial laser scanning. *Natural Hazards and Earth System Sciences* 9(3): 1003–1019.

Oskin ME, et al. (2012) Near-field deformation from the El Mayor–Cucapah earthquake revealed by differential LIDAR. *Science* 335(6069): 702–705.

Phillips DA, Jackson ME, and Meertens C (2008) GeoEarthScope airborne LiDAR and satellite InSAR imagery. *EOS Transactions AGU* 89: 53, abstract G53A-0630.

Pierrot-Deseilligny M and Paparoditis N (2006) A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. In: *ISPRS Workshop on Topographic Mapping from Space (With Special Emphasis on Small Satellites)* Ankara: IAPRS.

Prentice CS, Crosby CJ, Harding DJ, et al. (2003) Northern California LiDAR data: A tool for mapping the San Andreas fault and Pleistocene marine terraces in heavily vegetated terrain. *EOS Transactions, AGU* 84: 46, abstract G12A-06.

Reffet E, et al. (2010) Formation and stability of transverse and longitudinal sand dunes. *Geology* 38(6): 491–494.

Richardson MI, et al. (2007) PlanetWRF: A general purpose, local to global numerical model for planetary atmospheric and climate dynamics. *Journal of Geophysical Research-Planets* 112(E9).

Roering JJ, et al. (2009) Using DInSAR, airborne LiDAR, and archival air photos to quantify landsliding and sediment transport. *Geophysical Research Letters* 36.

Sabater N, et al. (2012) Contrast invariant and affine sub-pixel optical flow. In: *19th IEEE International Conference on Image Processing (ICIP),* pp. 53–56.

Scambos TA, et al. (1992) Application of image cross-correlation to the measurement of glacier velocity using satellite image data. *Remote Sensing of Environment* 42(3): 177–186.

Scharstein D and Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47(1–3): 7–42.

Scherler D, et al. (2008) Glacier-surface velocities in alpine terrain from optical satellite imagery—Accuracy improvement and quality assessment. *Remote Sensing of Environment* 112(10): 3806–3819.

Scherler D, et al. (2011) Spatially variable response of Himalayan glaciers to climate change affected by debris cover. *Nature Geoscience* 4(3): 156–159.

Shan S, et al. (2007) Kinematic GPS solutions for aircraft trajectories: Identifying and minimizing systematic height errors associated with atmospheric propagation delays. *Geophysical Research Letters* 34(23).

Sieh K, et al. (1993) Near-field investigations of the Landers earthquake sequence April to July 1992. *Science* 260: 171–176.

Singhvi AK and Porat N (2008) Impact of luminescence dating on geomorphological and palaeoclimate research in drylands. *Boreas* 37(4): 536–558.

Slatton KC, et al. (2007) Airborne laser swath mapping: Achieving the resolution and accuracy required for geosurficial research. *Geophysical Research Letters* 34(23).

Squarzoni C, et al. (2003) Nine years of spatial and temporal evolution of the La Valette landslide observed by SAR interferometry. *Engineering Geology* 68(1–2): 53–66.

Sun D, et al. (2010) Secrets of optical flow estimation and their principles. In: *2010 IEEE Conference on Computer Vision and Pattern Recognition,* pp. 2432–2439.

Tarolli P, et al. (2009) Understanding earth surface processes from remotely sensed digital terrain models. *Geomorphology* 113(1–2): 1–3.

Taylor MH, et al. (2008) Detecting co-seismic displacements in glaciated regions: An example from the great November 2002 Denali earthquake using SPOT horizontal offsets. *Earth and Planetary Science Letters* 270(3–4): 209–220.

Teza G, et al. (2007) Terrestrial laser scanner to detect landslide displacement fields: A new approach. *International Journal of Remote Sensing* 28(16): 3425–3446.

Titus TN, et al. (2008) Priorities for future research on planetary dunes. *Eos, Transactions American Geophysical Union* 89(45): 447–448.

Travelletti J, et al. (2012) Correlation of multi-temporal ground-based optical images for landslide monitoring: Application, potential and limitations. *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 39–55.

Treiman J, et al. (2002) Primary surface rupture associated with the M 7.1 16 October 1999 Hector Mine Earthquake, San bernardino County, California. *Bulletin of the Seismological Society of America* 92(4): 1171–1191.

Vadon H and Massonnet D (2000) Earthquake displacement fields mapped by very precise correlation. Complementarity with radar interferometry, Igarss 2000. In: *IEEE 2000 International Geoscience and Remote Sensing Symposium, Vol. I–Vi, Proceedings,* pp. 2700–2702.

Van Puymbroeck N, et al. (2000) Measuring earthquakes from optical satellite images. *Applied Optics* 39(20): 3486–3494. http://dx.doi.org/10.1029/2008gl035921.

Vermeesch P and Drake N (2008) Remotely sensed dune celerity and sand flux measurements of the world's fastest barchans (Bodele, Chad). *Geophysical Research Letters* 35(24). http://dx.doi.org/10.1029/2008g/035921.

Vermeesch P, et al. (2010) Sand residence times of one million years in the Namib Sand Sea from cosmogenic nuclides. *Nature Geoscience* 3(12): 862–865.

Wei SJ, et al. (2011) Superficial simplicity of the 2010 El Mayor-Cucapah earthquake of Baja California in Mexico. *Nature Geoscience* 4(9): 615–618.

Willert CE and Gharib M (1991) Digital image particle image velocimetry. *Experiments in Fluids* 10(4): 181–193.

Wolf PR and Dewitt BA (2000) *Elements of Photogrammetry with Application in GIS*, 3rd edn. India: McGraw-Hill Education.

Woolard JW and Colby JD (2002) Spatial characterization, resolution, and volumetric change of coastal dunes using airborne LIDAR: Cape Hatteras, North Carolina. *Geomorphology* 48(1–3): 269–287.

Yeats RS, et al. (1996) *The Geology of Earthquakes,* p. 568. New York: Oxford University Press.

Yu S-B, et al. (2001) Preseismic deformation and coseismic displacements associated with the 1999 Chi-Chi, Taiwan, Earthquake. *Bulletin of the Seismological Society of America* 91(5): 995–1012.

Zabih R and Woodfill J (1994) *Paper Presented at 3rd European Conf. Computer Vision, Stockholm.*

Zhang ZY (1994) In: Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision* 13(2): 119–152.

Zielke O, et al. (2010) Slip in the 1857 and earlier large earthquakes along the Carrizo Plain, San Andreas Fault. *Science* 327(5969): 1119–1122.

Zielke O, et al. (2012) High-resolution topography-derived offsets along the 1857 Fort Tejon Earthquake Rupture Trace, San Andreas Fault. *Bulletin of the Seismological Society of America* 102(3): 1135–1154.

Zitova B and Flusser J (2003) Image registration methods: A survey. *Image and Vision Computing* 21(11): 977–1000.